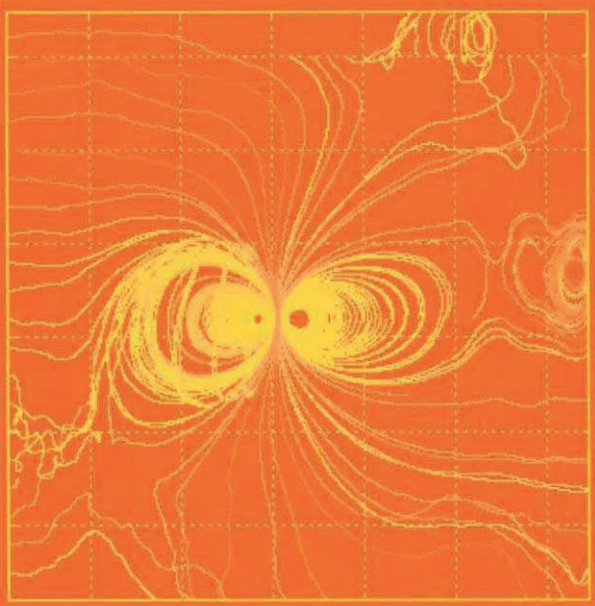
LECTURE NOTES IN PHYSICS

J. Büchner C.T. Dum M. Scholer (Eds.)

Space Plasma Simulation



Springer

Lecture Notes in Physics

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 Each contribution should contain at least partly original and previously unpublished material, be written in a clear, pedagogical style and aimed at a broader readership, especially graduate students and nonspecialist researchers wishing to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/lnp/

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

J. Büchner C. T. Dum M. Scholer (Eds.)

Space Plasma Simulation

Springer

Editors

Jörg Büchner
Max-Plack-Institut für Aeronomie,
Max-Planck Str. 2,
37191 Katlenburg-Lindau, Germany

Christian T. Dum
Max-Planck-Institut
für extraterrestrische Physik,
Postfach 1312
85741 Garching, Germany

Manfred Scholer
Max-Planck-Institut
für extraterrestrische Physik,
Postfach 1312
85741 Garching, Germany

Cover Picture: (see contribution by D. Cai et al. in this volume)

Cataloging-in-Publication Data applied for

A catalog record for this book is available from the Library of Congress.

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at http://dnb.ddb.de

ISSN 0075-8450
ISBN 3-540-00698-2 Springer-Verlag Berlin Heidelberg New York

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

http://www.springer.de

Printed in Germany

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
54/3141/du - 5 4 3 2 1 0

Preface

The aim of this book is twofold: to provide an introduction for newcomers to state of the art computer simulation techniques in space plasma physics and an overview of current developments. Computer simulation has reached a stage where it can be a highly useful tool for guiding theory and for making predictions of space plasma phenomena, ranging from microscopic to global scales.

The various articles are arranged, as much as possible, according to the underlying simulation technique, starting with the technique that makes the least number of assumptions: a fully kinetic approach which solves the coupled set of Maxwell's equations for the electromagnetic field and the equations of motion for a very large number of charged particles (electrons and ions) in this field. Clearly, this is also the computationally most demanding model. Therefore, even with present day high performance computers, it is the most restrictive in terms of the space and time domain and the range of particle parameters that can be covered by the simulation experiments.

It still makes sense, therefore, to also use models, which due to their simplifying assumptions, seem less realistic, although the effect of these assumptions on the outcome of the simulation experiments needs to be carefully assessed. In fact, using a model which is not realistic in every respect may, instead of a limitation, even represent a particular strength of simulation. Such models allow isolating particular physical mechanisms, something which is often not possible with space observations or even with laboratory experiments, in view of their complexity. In this manner our theoretical understanding can be advanced.

One such simplification, often also employed in theoretical treatments, uses of the fact that plasma waves are often nearly electrostatic. Corresponding codes are not only simpler, because there is no need to compute magnetic field perturbations. They are also more efficient, because wave phase speeds that need to be resolved are generally much smaller than for electromagnetic waves. By comparing electrostatic and electromagnetic code results, one can easily assess the importance of electromagnetic effects in plasma dynamics.

Space plasmas, in contrast to laboratory devices, are very extended coupled systems. Boundary conditions, which a computation necessarily has to introduce, are thus always somewhat artificial. Using two- or even one-dimensional models allows minimizing this artifact in the remaining dimensions. Especially with one-dimensional models, however, careful analysis must ensure that they do not too much restrict interaction mechanisms, between particles and waves, for example.

Even the fully kinetic simulations by the particle in cell (PIC) simulation technique simplifies the computational task by concentrating on collective, long range electromagnetic fields, which dominate the dynamics of most space plasmas. This method of representing the plasma by finite size particles still leaves background statistical noise, with effects similar to particle collisions. It is reduced only as $N^{-\frac{1}{2}}$, as the number of such super particles increases. A method to more efficiently reduce the noise level is to split the particle distribution function, when this seems possible, into a fixed background f_0 and a perturbation δf and representing only the latter by simulation particles.

In a dusty plasma, however, the short range forces, which describe collisions, become important, at least between dust grains. The evaluation of these forces for N particles requires the summation of N^2 interactions. Moreover, dust grains are so much more massive than plasma ions and electrons, and hence move on vastly different time scales, which also largely prohibits using the same simulation technique as for the background plasma. One can, however, take advantage of this disparity in characteristic scales, by combining molecular dynamics simulation techniques for the dust grains with a description, obtained from linear kinetic theory, of the plasma as a dielectric which modifies the forces on dust grains.

A very significant reduction in computational cost arises if at least one particle species can reasonably be represented by a simple fluid model. In so called hybrid models, usually the electrons are represented in this way, whereas for the ions a fully kinetic description by particles is used. Situations in which, by contrast, electron kinetic effects dominate are also possible, of course.

The most significant reduction in computational cost arises if all particle species are represented by fluids, or even a single conducting fluid. In return, such magnetohydrodynamic models nowadays allow global simulations of the entire coupled solar wind-magnetospheric system. The ongoing challenge, as computer power increases, is to make the description of the plasma and the boundary conditions, especially the planetary boundary conditions, more complete. Predictions of space weather require as input, at another boundary, also realistic descriptions of solar activity.

Along with advances in computer hardware, advanced programming and numerical techniques, which make optimal use of this new computer power, such as massively parallel computations or codes with variable mesh size, are equally important. Finally, special attention must be paid to efficient techniques for analyzing and representing the massive amount of data, which simulations, three-dimensional computations especially, nowadays produce.

This book was inspired by the Sixth International School / Symposium for Space Plasma Simulation, held at Garching, Germany, in September 2001. Its invited tutorials and review lectures had the same aims. We therefore invited authors to contribute who, in a collective effort, would give a most coherent and reasonably complete picture of this diverse and rich field. To this end individual contributions were made available to all contributing authors during the preparation phase of this book, and authors were encouraged to freely exchange comments. In addition, each contribution was subject to a formal peer review

process. We feel very fortunate to have found such highly qualified authors who, as a group, have indeed covered a huge range of topics, as the table of contents shows. Our thanks go to them, but also to the referees who helped to further improve the presentation.

Garching and Katlenburg-Lindau
December 2002 J. Büchner, C. T. Dum, M. Scholer

Table of Contents

List of Contributors

Dongsheng Cai
Institute of Information Sciences and Electronics
The University of Tsukuba,
Ibaraki 305-8573,
Japan
cai@is.tsukuba.ac.jp

Darren L. De Zeeuw
Center for Space Environment Modeling
The University of Michigan,
Ann Arbor, MI 48109-2143
USA

Thierry Dudok de Wit
LCPE, UMR 6115 CNRS-Université d'Orléans,
3A, av. de la Recherche Scientifique,
Orléans cedex 2, 45071
France

Gurudas Ganguli
Plasma Physics Division
Naval Research Laboratory
Washington, DC 20375,
USA

Tamas I. Gombosi
Center for Space Environment Modeling
The University of Michigan,
Ann Arbor, MI 48109-2143
USA

Joseph D. Huba
Plasma Physics Division
Naval Research Laboratory
Washington, DC 20375,
USA

Glenn Joyce
Plasma Physics Division
Naval Research Laboratory
Washington, DC 20375,
USA

Homa Karimabadi
Electrical and Computer Engineering Department
University of California San Diego
La Jolla, CA 92093,
USA

Martin Lampe
Plasma Physics Division
Naval Research Laboratory
Washington, DC 20375,
USA

Bertrand Lembège
CETP-CNRS-UVSQ
10-12 Av. de l'Europe
Velizy, 78140
France

Yaoting Li
Institute of Information Sciences and Electronics
The University of Tsukuba,
Ibaraki 305-8573,
Japan
ytli@is.tsukuba.ac.jp

Hiroshi Matsumoto
Radio Science Center for Space and Atmosphere
Kyoto University
Uji, Kyoto 611-001,
Japan

Ken-Ichi Nishikawa
National Space Science and Technology Center,
320 Sparkman Drive, SD 50,
Huntsville, AL 35805,
USA
Ken-Ichi.Nishikawa@nsstc.nasa.gov

Tatsuki Ogino
Solar-Terrestrial Environment Laboratory,
Nagoya University 3-13 Honohara,
Toyokawa, Aichi 442-8507,
Japan
ogino@stelab.nagoya-u.ac.jp

Nick Omidi
Electrical and Computer Engineering Department
University of California San Diego
La Jolla, CA 92093,
USA
nomidi1@san.rr.com

Yoshiharu Omura
Radio Science Center for Space and Atmosphere
Kyoto University
Uji, Kyoto 611-001,
Japan

Antonius Otto
Geophysical Institute
University of Alaska
Fairbanks, AK 99775,
USA

Kenneth G. Powell
Center for Space Environment Modeling
The University of Michigan,
Ann Arbor, MI 48109-2143
USA

Philip L. Pritchett
Department of Physics and Astronomy
University of California Los Angeles
Los Angeles, CA 90095-1547,
USA
pritchet@physics.ucla.edu

Zuying Pu
Department of geophysics,
Peking University,
Beijing 100871,
China
zypu@pku.edu.cn

Kevin Quest
Electrical and Computer Engineering Department
University of California San Diego
La Jolla, CA 92093,
USA

Joachim Raeder
Institute of Geophysics and Planetary Physics
University of California Los Angeles
Los Angeles, 90095-1567
USA

Aaron J. Ridley
Center for Space Environment Modeling
The University of Michigan,
Ann Arbor, MI 48109-2143
USA

David Schriver
University California Los Angeles
Los Angeles, CA 90095-1567,
USA
dave@igpp.ucla.edu

Igor V. Sokolov
Center for Space Environment Modeling
The University of Michigan,
Ann Arbor, MI 49109-2143
USA

Quentin F. Stout
Center for Space Environment Modeling
The University of Michigan,
Ann Arbor, MI 49109-2143
USA

Richard D. Sydora
Department of Physics
University of Alberta
Edmonton, AB, T6G 2J1
Canada
rsydora@phys.ualberta.ca

Takashi Tanaka
Communications Research Laboratory
Koganei-shi, Tokyo 184-8795,
Japan

Gábor Tóth
Center for Space Environment Modeling
The University of Michigan,
Ann Arbor, MI 48109-2143
USA

Takayuki Umeda
Radio Science Center for Space and Atmosphere
Kyoto University
Uji, Kyoto 611-001,
Japan

Dan Winske
Applied Physics Division
Los Alamos National Laboratory
Los Alamos, NM 87545,
USA
winske@lanl.gov

Chijie Xiao
Institute of Information Sciences and Electronics
The University of Tsukuba,
Ibaraki 305-8573,
Japan
cjxiao@is.tsukuba.ac.jp

Xiaoyang Yan
Institute of Information Sciences and Electronics
The University of Tsukuba,
Ibaraki 305-8573,
Japan
yxy@is.tsukuba.ac.jp

Lin Yin
Applied Physics Division
Los Alamos National Laboratory
Los Alamos, NM 87545,
USA
lyin@lanl.gov

Hua Zhu
Geophysical Institute
University of Alaska
Fairbanks, AK 99775,
USA

Particle-in-Cell Simulation of Plasmas – A Tutorial

Philip L. Pritchett

Department of Physics and Astronomy, University of California, Los Angeles, CA 90095-1547, USA

Abstract. This chapter serves as a tutorial introduction to the field of particle-in-cell (PIC) simulation of plasmas. The full particle version of these models, in which both the electrons and ions are treated as particles, makes no approximations to the basic laws of mechanics and electricity and magnetism, and thus the full range of collisionless plasma physics is included in such a model. The basic techniques involved in a PIC model are illustrated: accumulation of the charge and current densities on a spatial grid, time integration of the field and particle equations, and limitations imposed by the underlying kinetic physics of a plasma. Various approximations to the full set of Maxwell's equations are described as well as the case of hybrid models (particle ions, fluid electrons) in which fluid equations are used to model the slow time evolution of high frequency phenomena, thus allowing a study of lower frequency kinetic phenomena on longer temporal and larger spatial scales. Examples of PIC simulations of magnetic reconnection are discussed.

1 Introduction

Plasma is frequently characterized as the fourth state of matter. It consists of electrons, ions, and neutral atoms, usually at temperatures above 10^4 Kelvin. Plasma is pervasive throughout the universe: from stars to the sun to the Earth's magnetosphere to terrestrial experiments attempting to harness the power of thermonuclear fusion. As in most fields of science, the traditional approaches to studying the properties of plasmas have involved experiment and observation on the one hand and analytical techniques based on fundamental physical laws on the other. Progress comes from an interplay between these approaches: one probes nature through experiments, and the results are used to confirm or disprove the theoretical expectations.

The explosive growth in the power of computers over the past half century has led to the development of a third approach to the study of science in general and plasma physics in particular: computational physics. The impact of this alternative approach has been particularly prominent in plasma physics. The basic laws governing plasma behavior, namely the laws of Newton and Maxwell, are well known, but the consequences of these laws for a complex system consisting of perhaps 10^{20} or more particles are frequently impossible to determine. Likewise, the necessary experiments may be difficult and expensive to perform, either because the plasma is located far away and must thus be probed via spacecraft or because one must construct a large device to achieve the desired plasma con-

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 1–24, 2003.

ditions in a laboratory setting. Given these difficulties, computer simulation of plasmas has frequently proven to be a more practical means to make progress.

Traditionally, there have been two complementary approaches to computational plasma physics. One emphasizes the fluid nature of plasma and proceeds by solving numerically the magnetohydrodynamic equations of a plasma assuming approximate transport coefficients. The other gives primacy to the kinetic interactions among the constituent particles of the plasma and the electromagnetic field. The fluid approach is more amenable to treating large scale properties of plasmas involving mass, momentum, and flux transport, while the kinetic description provides a more accurate treatment of many local and quasi-local processes. This chapter will be concerned with only the kinetic approach to plasma simulation.

The usual basis for analytic treatments of a collisionless plasma is the Vlasov equation:

$$\frac{\partial f_j}{\partial t} + \boldsymbol{v} \cdot \frac{\partial f_j}{\partial \boldsymbol{x}} + \frac{q_j}{m_j}\left(\boldsymbol{E} + \frac{\boldsymbol{v} \times \boldsymbol{B}}{c}\right) \cdot \frac{\partial f_j}{\partial \boldsymbol{v}} = 0 \,. \tag{1}$$

Here, $f_j(\boldsymbol{x}, \boldsymbol{v}, t)$ is the distribution function of the jth species, and the electric and magnetic fields are determined by Maxwell's equations. The Vlasov equation represents a partial differential equation in a six-dimensional phase space plus time. To fully resolve this entire phase space requires an amount of computer memory that exceeds what is available on today's biggest machines. For example, to have 30/30/30 point resolution of velocity space and 128/128/128 in configuration space would require approximately 10^3 Gbytes. While numerical models have been developed for solving the Vlasov-Maxwell equations directly (e.g., [43], [44]), they are typically restricted to two or three phase space dimensions and frequently assume a fixed magnetic field. We shall not consider this direct approach further.

The more common approach to kinetic modeling of plasma is to represent f_j by a number of macroparticles and to compute the particle orbits in the self-consistent electric and magnetic fields. This is equivalent to solving the Vlasov equation by the method of characteristics. The early models (e.g., [17]) treated the particles as discrete points and computed the electric force acting on each particle by summing explicitly the Coulomb interaction with each of the other $N-1$ particles. The number of pairs of such interactions is given by $N(N-1)/2$. This N^2 variation is a very restrictive limitation on the number of particles that can be employed. For example, suppose that we consider a calculation involving 10^7 particles (a medium size run by today's standards) extending over 10^4 time steps. If the evaluation of the force is estimated to involve 10 arithmetic operations, then the total number of operations $\sim 5 \times 10^{18}$. If each operation requires $\sim 10^{-9}$ s, then the total time $\sim 5 \times 10^9$ s ≈ 160 years. This is clearly not practical.

The solution which was developed and has now become standard is to introduce a spatial grid on which the particles' charge and current densities are accumulated using an interpolation scheme. The field equations (all or some of Maxwell's equations) are solved on this grid, and the forces acting on the

particles are obtained by interpolating the fields back to the particles. This is the "particle-in-cell" (PIC) technique. This procedure eliminates fluctuations at scales smaller than the grid spacing and also reduces the number of operations per time step to $\sim N \log N$. To be internally consistent, PIC models must deal with situations in which any fields are slowly varying on the scale of the grid spacing so that the normal modes of the plasma are properly resolved. We will be exclusively concerned with PIC models in this chapter. There is a very extensive and detailed literature on the subject, including review articles [18], [25], the monographs *Methods in Computational Physics* [1], [2], textbooks [5], [37], [82], and the collections of papers from the previous International Schools for Space Simulation [3], [54], [63], [64]. There are, however, instances where certain aspects of plasma systems involving short range forces are effectively treated by the direct N particle or "molecular dynamics" approach [37]. An example is the case of dusty plasmas discussed in the chapter by *Joyce et al.* [40] in this volume.

The outline of the chapter is as follows. In Sect. 2 we discuss the full particle version of a PIC simulation in terms of a model that solves the full set of Maxwell's equations. Consideration is given to the accumulation of the particle source terms on the grid, the time advancement of the field and particle equations, and consequences of the spatial grid. The modifications introduced by various approximations to Maxwell's equations are discussed as is the role of implicit vs. explicit models. Section 3 discusses hybrid simulations in which the electrons are described by fluid equations. Some examples of full particle simulations are presented in Sect. 4. Section 5 contains a final discussion.

2 Full Particle Models

In a full particle model one follows the motion of both electrons and ions in the self-consistent electric and magnetic fields obtained from a solution of Maxwell's equations. Relativistic effects are readily included by the use of the Lorentz equations of motion for the particles. At this level one has introduced no approximations in the basic laws of mechanics and electricity and magnetism, and thus the full range of collisionless plasma physics is included in such a model. Such a model represents the closest approach to mimicing real plasma behavior of all simulation models. As we shall see, this fidelity to nature also is the source of many limitations.

2.1 Electromagnetic Models

Models which solve the full set of Maxwell's equations are termed "electromagnetic." One has the choice of working directly with the fields $\boldsymbol{E}$ and $\boldsymbol{B}$, in which case the equations are

$$\partial \boldsymbol{E}/\partial t = c(\boldsymbol{\nabla} \times \boldsymbol{B}) - 4\pi \boldsymbol{J} \,, \tag{2}$$

$$\partial \boldsymbol{B}/\partial t = -c(\boldsymbol{\nabla} \times \boldsymbol{E}) \,, \tag{3}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{E} = 4\pi\varrho \ , \quad \boldsymbol{\nabla} \cdot \boldsymbol{B} = 0 \ , \tag{4}$$

or introducing the vector and scalar potentials $\boldsymbol{A}$ and Φ, in which case the field equations (in the Coulomb gauge) are

$$\partial^2 \boldsymbol{A}/\partial t^2 = c^2 \nabla^2 \boldsymbol{A} + 4\pi c \boldsymbol{J}_{\mathrm{T}} \ , \tag{5}$$

$$\nabla^2 \Phi = -4\pi\varrho \ , \tag{6}$$

$$\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A} \ , \quad \boldsymbol{E} = -(1/c)\partial \boldsymbol{A}/\partial t - \boldsymbol{\nabla}\Phi \ . \tag{7}$$

Here, the transverse current $\boldsymbol{J}_{\mathrm{T}}$ is given by

$$\boldsymbol{J}_{\mathrm{T}} = \boldsymbol{J} - (1/4\pi)\boldsymbol{\nabla}(\partial\Phi/\partial t) \ . \tag{8}$$

We shall primarily discuss the $\boldsymbol{E}$,$\boldsymbol{B}$ version which is slightly simpler numerically; comparison with theory, however, is sometimes more convenient in terms of $\boldsymbol{A}$,Φ. There appears to be no fundamental numerical advantage in favor of one or the other [50].

Particle Source Terms

The first step in solving Maxwell's equations is to determine the source terms $\boldsymbol{J}$ and ϱ. As we discussed in the Introduction, the essence of the PIC approach is that the current and charge densities are accumulated on a spatial grid from the particle data using an interpolation procedure. The earliest PIC codes used the nearest grid point approximation, but this approach resulted in excessive fluctuations as particles moved across the grid. Most codes now use a linear interpolation scheme, which involves 2 points in 1D, 4 points in 2D, and 8 points in 3D. For example, given a 1-D spatial grid with uniform spacing Δ_g, the charge q_j of a particle located at x_j makes contributions to the charge density at 2 grid points as follows:

$$\begin{aligned} \varrho(x_{\mathrm{L}}) \quad &= q_j(1 - \Delta x_j/\Delta_g) \ , \\ \varrho(x_{\mathrm{L}}+1) &= q_j \Delta x_j/\Delta_g \ , \end{aligned} \tag{9}$$

where x_{L} is the grid point immediately to the left of the particle and $\Delta x_j \equiv x_j - x_{\mathrm{L}}$. Thus the charge is distributed between the 2 nearest grid points. The total charge density on the grid is then obtained by summing over all the particles. A similar procedure is used for the current density.

Time Integration

In analytic treatments of Maxwell's equations, it is sufficient to solve the time-dependent equations (2) and (3). If the fields satisfy the divergence equations (4) initially, they will at all subsequent times as well. The situation is more complicated for PIC models, but the time-dependent equations still receive dominant attention. One normally solves (2) and (3) by introducing a leapfrog scheme as

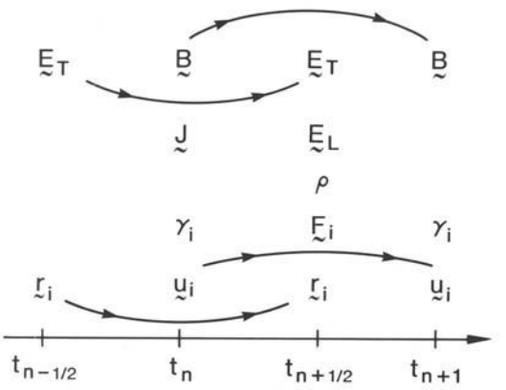


Fig. 1. The time sequencing of particle and field quantities involved in advancing the electromagnetic simulation code through one cycle

illustrated in Fig. 1. $\boldsymbol{E}$ is defined at half-integral time steps $(n-1/2)\Delta t$ and $\boldsymbol{B}$ at integral time steps $n\Delta t$. One then uses a finite difference approximation to the time derivatives to advance the fields through a time step Δt:

$$\boldsymbol{E}^{n+1/2} = \boldsymbol{E}^{n-1/2} + \Delta t[c(\boldsymbol{\nabla} \times \boldsymbol{B}^n) - 4\pi \boldsymbol{J}^n] \,, \tag{10}$$

$$\boldsymbol{B}^{n+1} = \boldsymbol{B}^n - c\Delta t \boldsymbol{\nabla} \times \boldsymbol{E}^{n+1/2} \,. \tag{11}$$

Notice that the curls and the current density on the right hand side of (10) and (11) are evaluated at a time midway between the old and new values of the fields. Such a scheme is said to be "time-centered," and the integration is then accurate to second order in Δt. The relativistic particle equations of motion are likewise integrated in time using a leapfrog scheme, with the particle coordinates at the half-integral times and the momenta at the integral times. One cycle of the model then involves the advancement from $\{\boldsymbol{r}_i^{n-1/2}, \boldsymbol{v}_i^n, \boldsymbol{E}^{n-1/2}, \boldsymbol{B}^n\}$ to $\{\boldsymbol{r}_i^{n+1/2}, \boldsymbol{v}_i^{n+1}, \boldsymbol{E}^{n+1/2}, \boldsymbol{B}^{n+1}\}$. (When the relativistic force equations are used, one advances the momentum per unit mass $\boldsymbol{u} = \boldsymbol{p}/m_0$ instead of the velocity $\boldsymbol{v}$. The Lorentz gamma factor is given by $\gamma = (1 + u^2/c^2)^{1/2}$.)

What determines the choice of Δt? As an example, let us apply (10) and (11) to the case of electromagnetic waves in vacuum ($\boldsymbol{J} = 0$). We will also assume that the spatial dependence of $\boldsymbol{E}$ and $\boldsymbol{B}$ can be expressed in terms of Fourier amplitudes such that

$$\boldsymbol{E}(\boldsymbol{k}, t) = \boldsymbol{E}_0(\boldsymbol{k}) \exp[-\mathrm{i}\omega(n-1/2)\Delta t] \,, \tag{12}$$

$$\boldsymbol{B}(\boldsymbol{k}, t) = \boldsymbol{B}_0(\boldsymbol{k}) \exp[-\mathrm{i}\omega n \Delta t] \,. \tag{13}$$

Substituting (12) and (13) into (10) and (11), we obtain

$$\sin^2(\omega \Delta t/2) = k^2 c^2 (\Delta t)^2/4 \,. \tag{14}$$

This is the dispersion relation that determines the wave frequency ω in terms of the wavenumber k. As $\Delta t \to 0$, we recover the expected relation $\omega = \pm kc$. For finite Δt, we see that there are real solutions for ω only if $k^2c^2(\Delta t)^2 < 4$. For larger values of Δt, there are only complex conjugate solutions for ω, and the solutions (12) and (13) will then grow exponentially in time. The leapfrog integration scheme is then unstable. This is an example of a general result known as the *Courant-Friedrichs-Lewy* [13] condition: for an explicit time integration scheme the time step is limited by the largest k mode or highest frequency oscillation that enters the problem.

In addition to being stable, we want the integration scheme to be accurate. Thus we are also interested in the magnitude of the frequency error given by (14) relative to the exact result $\omega_0 = \pm kc$. For small $kc\Delta t$, the solution of (14) is

$$\omega = \pm kc[1 + k^2c^2(\Delta t)^2/24 + \cdots \ ] \ . \tag{15}$$

The second term in brackets represents the phase error introduced by the finite time step. After a time $T = N\Delta t$, the accumulated phase error for the oscillation will be $(\omega - \omega_0)T = \omega_0 T(\omega_0 \Delta t)^2/24$. Thus decreasing Δt will lower the error at the expense of increasing the number of steps N.

Another important time step limitation arises from oscillations at the plasma frequency $\omega_{pe} = (4\pi n_0 e^2/\ m_e)^{1/2}$. Using the same analysis as for the electromagnetic wave example above, one finds that the leapfrog algorithm applied to a simple harmonic oscillator with frequency ω_{pe} becomes unstable when $\omega_{pe}\Delta t > 2$. Simulations with a thermal plasma [47] indicate that the instability threshold is reduced to $\omega_{pe}\Delta t > 1.62$. For accurate reproduction of the plasma oscillations, one needs to choose a considerably smaller Δt; a typical choice is $\omega_{pe}\Delta t \sim$ 0.1–0.2.

Spatial Grid

We have not yet dealt with the representation of the spatial derivatives in Maxwell's equations. Two different approaches have been used. One is to introduce a finite difference approximation, exactly in analogy with the case of the time derivatives. For reasons of accuracy, it is again desirable to employ a centered difference approximation. This means that the spatial grids for $\boldsymbol{E}$ and $\boldsymbol{B}$ should be displaced relative to each other. This is normally done based on the *Yee* [92] lattice, which is a fully staggered grid mesh system. The components of $\boldsymbol{E}$ and $\boldsymbol{J}$ are defined at midpoints of cell edges, while the components of $\boldsymbol{B}$ are defined at the midpoints of the cell surfaces (see Fig. 1 of [88]). With such a scheme, $\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0$ will be maintained to machine roundoff. Poisson's equation, however, requires special attention. The problem arises from the fact that the grid quantities $\boldsymbol{J}$ and ϱ produced by most simple interpolation schemes may not satisfy the continuity equation,

$$\partial\varrho/\partial t + \boldsymbol{\nabla} \cdot \boldsymbol{J} = 0 \ . \tag{16}$$

If this equation is violated, then the straightforward integration of (2) and (3) will rapidly lead to the development of nonphysical fields which do not satisfy $\boldsymbol{\nabla} \cdot$

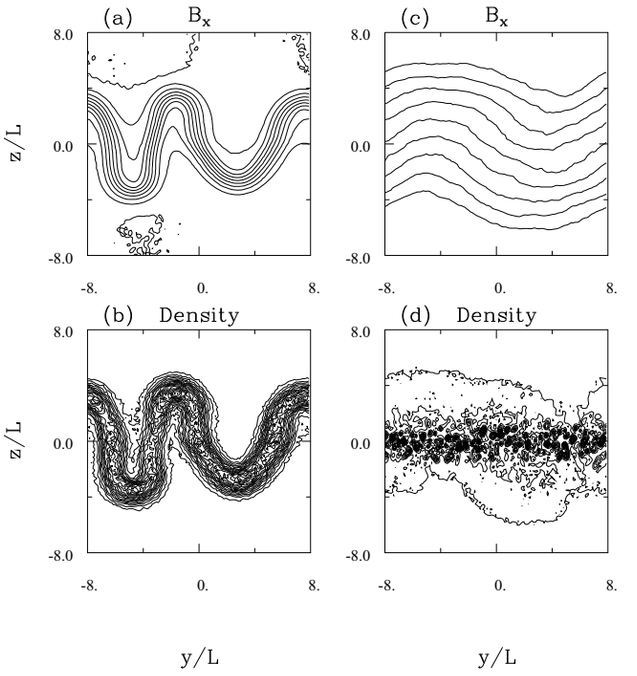


Fig. 2. Contour plots in the y,z plane of the magnetic field B_x and of the density at time $\Omega t = 60$. The left panels (**a**) and (**b**) are from a simulation with a full Poisson correction for the electric field, while the right panels (**c**) and (**d**) are from a run with no correction

$\boldsymbol{E} = 4\pi\varrho$. Thus, one must either adopt an interpolation scheme which rigorously satisfies the continuity equation (e.g., [87], [88]) or else make a correction to ensure that $\boldsymbol{\nabla} \cdot \boldsymbol{E} = 4\pi\varrho$ is maintained (e.g., [50]). This second scheme, which is much simpler to implement, involves adding a correction $\delta\boldsymbol{E}$ to the electric field computed from (2) which is determined by solving $\nabla^2(\delta\Phi) = -(4\pi\varrho - \boldsymbol{\nabla} \cdot \boldsymbol{E})$ and setting $\delta\boldsymbol{E} = -\boldsymbol{\nabla}\delta\Phi$.

Figures 2–4 illustrate the consequences of failing to ensure that the continuity equation is satisfied. These results are from a 2-D simulation in the y,z plane starting from the simple Harris neutral sheet with $B_{0x}(z) = B_0 \tanh(z/L)$. The mass ratio is $m_i/m_e = 1$, the current sheet thickness is $\rho_0/L = 0.5$ (ρ_0 is the particle gyroradius in the B_0 field), and the grid size is $N_y \times N_z = 128 \times 128$. In this configuration the tearing mode instability cannot occur, but finite-k_y instabilities such as the drift kink mode [67], [70], [94] are allowed. Figures 2a and 2b show contour plots of the magnetic field B_x and the density at time $\Omega t = 60$ obtained in a run with a full Poisson correction for the electric field. A strong $k_y = 2$ kink mode (wavenumber $k_y L = 0.79$) is quite apparent. In contrast, Figs. 2c and 2d show the corresponding plots where no correction was applied to the electric field. It is clear that most of the kink structure has been

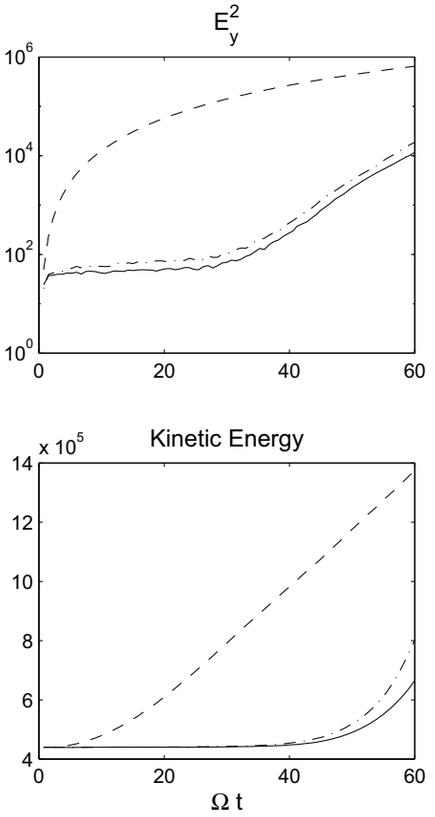


Fig. 3. Time histories of the total E_y^2 field energy and the total kinetic energy in 2-D y,z simulations with a full Poisson correction for the electric field (*solid lines*), no correction (*dashed lines*) and a single-pass Point-Jacobi correction (*dash-dot lines*)

washed away and the density is dominated by non-physical short wavelength noise.

Figure 3 shows the time history of the total E_y^2 field energy and of the kinetic energy. The solid lines are for the run with the Poisson correction, while the dashed lines are for the case with no correction. This latter case exhibits unphysical energy growth. The dash-dot lines are for a run employing a single-pass point-Jacobi correction rather than a full Poisson solution [48]. In this approximation the correction potential $\delta\Phi = (\Delta_g^2/4)(4\pi\varrho - \boldsymbol{\nabla}\cdot\boldsymbol{E})$. This provides a significant improvement over the case of no correction. Figure 4 shows the time history of the A_z amplitude squared for mode 2. This demonstrates the failure of the uncorrected run to preserve the growth of the kink mode after $\Omega t \approx 30$.

The second approach is to obtain the solution of (2)–(4) in Fourier space. The transformations between coordinate space and Fourier space are carried out via fast Fourier transforms (FFT) (see Appendix A of [5].) To this end, it is useful

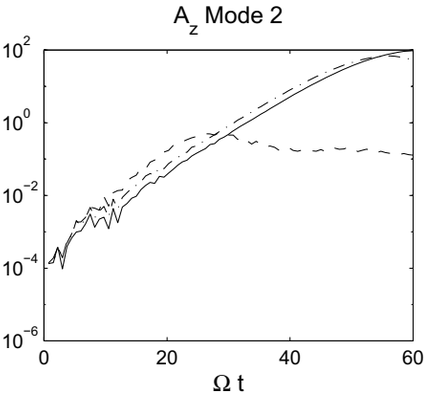


Fig. 4. Time history for the absolute value squared of the vector potential A_z for mode 2 for the same set of simulations as in Fig. 3

to introduce the decomposition of an arbitrary vector $\boldsymbol{S}$ into its divergence free (transverse) and curl free (longitudinal) parts:

$$\boldsymbol{k} \cdot \boldsymbol{S}_{\mathrm{T}}(\boldsymbol{k}) = 0 \ , \qquad \boldsymbol{k} \times \boldsymbol{S}_{\mathrm{L}}(\boldsymbol{k}) = 0 \ . \tag{17}$$

From (4) it is clear that the magnetic field has only a transverse component. The time-dependent Maxwell equations then become

$$\partial \boldsymbol{E}_{\mathrm{T}}(\boldsymbol{k})/\partial t = ic\boldsymbol{k} \times \boldsymbol{B}(\boldsymbol{k}) - 4\pi \boldsymbol{J}_{\mathrm{T}}(\boldsymbol{k}) \ , \tag{18}$$

$$\partial \boldsymbol{B}(\boldsymbol{k})/\partial t = -ic\boldsymbol{k} \times \boldsymbol{E}_{\mathrm{T}}(\boldsymbol{k}) \ . \tag{19}$$

Poisson's equation determines the longitudinal part of $\boldsymbol{E}$, $i\boldsymbol{k} \cdot \boldsymbol{E}_{\mathrm{L}}(\boldsymbol{k}) = 4\pi\varrho(\boldsymbol{k})$. Maxwell's equations thus reduce to a set of algebraic equations for the Fourier amplitudes. This Fourier solution is quite accurate, but it has the disadvantage of being directly applicable only to cases where the system is periodic in each spatial direction. Generalizations to cases involving combinations of periodic and bounded coordinates were considered in [20].

Just as the plasma frequency ω_{pe} imposes a restriction on the allowable time step, so the Debye length $\lambda_{De} = v_{Te}/\omega_{pe} = (kT_e/4\pi n_0 e^2)^{1/2}$ imposes a restriction on the value of the grid spacing Δ_g. The problem arises when λ_{De} becomes considerably smaller than Δ_g. The effect of the spatial grid is to couple plasma perturbations to perturbations at other wavelengths known as aliases. This effect becomes larger when $\lambda_{De} \ll \Delta_g$, and it can lead to the growth of a numerical instability that can cause heating of the initial Maxwellian plasma [46], [56]. A general rule of thumb is that this effect has ignorable consequences for $\lambda_{De}/\Delta_g \gtrsim 0.3$ for the case of linear interpolation to and from the grid. Thus typically λ_{De}/Δ_g is chosen to be ~ 0.5–1.0.

2.2 Approximations to Maxwell's Equations

In some cases, it may not be necessary or desirable to solve the full set of Maxwell's equations. A particularly simple approximation is to treat the magnetic field as a fixed external field. The only self-consistent field is then the electric field, which is obtained from Poisson's equation (6). Such a model is termed "electrostatic." It has the great advantage of eliminating the light waves and thus removes the CFL constraint involving the speed of light. A number of projects using a 1-D electrostatic PIC code are discussed in Chap. 5 of [5].

A scheme which retains self-consistent magnetic and inductive electric fields but neglects radiation is the Darwin or magnetoinductive approximation [14]; it is based on an expansion correct to order $(v/c)^2$ [39]. This scheme neglects the transverse part of the displacement current so that (2) reverts back to Ampère's law,

$$\boldsymbol{\nabla} \times \boldsymbol{B} = (4\pi/c)\boldsymbol{J}_{\mathrm{T}} \; . \tag{20}$$

This approximation alters the field equations from a hyperbolic set to an elliptical set in which there is no longer any retardation effect. The solution for the transverse part of $\boldsymbol{E}$ then involves a more complicated field equation obtained by taking the curl of (3) and using the particle equations of motion to evaluate $(\partial \boldsymbol{J}/\partial t)_{\mathrm{T}}$ [10], [32], [66]. While the Darwin approach has frequently been used in 2-D simulations where it was possible to describe the electromagnetic fields solely in terms of the out-of-plane vector potential component [21] (and also in hybrid models, see Sect. 3), it loses its advantage in 3D and has particular difficulty in specifying the boundary conditions for the field solution in nonperiodic geometries [33]. Many of the advantages of the Darwin scheme can be more easily achieved by choosing an implicit scheme.

2.3 Explicit vs. Implicit Schemes

In all the time advancement schemes that we have considered so far, the new field and particle values were calculated from field and particle values at previous times only. Such a scheme is known as an "explicit" scheme. In order to be stable, an explicit scheme must satisfy all the CFL constraints. In contrast, in an "implicit" scheme the solution of the new quantities involves knowledge of these quantities at the new time, thus forming a potentially very large system of coupled nonlinear equations. The advantage of an implicit scheme is that it will be stable (albeit perhaps inaccurate) for large time steps.

A very simple version of an implicit scheme actually appears in the leapfrog algorithm for advancing the particles. The acceleration of a particle is given in terms of the Lorentz force,

$$\mathrm{d}\boldsymbol{v}_j/\mathrm{d}t = (q_j/m_j)[\boldsymbol{E} + (\boldsymbol{v}_j \times \boldsymbol{B})/c] \; . \tag{21}$$

Since the force depends on $\boldsymbol{v}_j$ itself, the question arises as to how to evaluate the right hand side. Using the leapfrog time scheme illustrated in Fig. 1, we can

time center the force by introducing an average of the old and new velocities. Thus the finite difference analog to (21) is

$$\begin{aligned} \boldsymbol{v}_j^{n+1} = \boldsymbol{v}_j^n + (q_j \Delta t/m_j)[\boldsymbol{E}^{n+1/2} \\ +(1/2)(\boldsymbol{v}_j^n + \boldsymbol{v}_j^{n+1}) \times \boldsymbol{B}^{n+1/2}/c] \,, \end{aligned} \tag{22}$$

where $\boldsymbol{B}^{n+1/2} \equiv (\boldsymbol{B}^n + \boldsymbol{B}^{n+1})/2$. This is an implicit equation because $\boldsymbol{v}_j^{n+1}$ appears on the right hand side. In this case, (22) comprises a set of three coupled linear equations for the components of $\boldsymbol{v}_j^{n+1}$, and the solution can be obtained using standard linear algebra techniques. An interesting consequence of this implicit approach is that the particle's gyromotion imposes no limit on the time step. Although the orbit will not be followed accurately if $\Omega_j \Delta t \gtrsim 1$ ($\Omega_j = q_j B/m_j c$ is the cyclotron frequency), the positions and velocities will not blow up.

One can use a similar approach to formulate the time advance for a fully implicit PIC simulation (e.g., [49]). For simplicity, we consider only the electric field contribution to the acceleration. Then we write

$$\boldsymbol{x}^{n+3/2} = \boldsymbol{x}^{n+1/2} + \Delta t \boldsymbol{v}^{n+1} \,, \tag{23}$$

$$\begin{aligned} \boldsymbol{v}^{n+1} = \boldsymbol{v}^n + \Delta t(q/2m)[\bar{\boldsymbol{E}}^{n-1/2}(\boldsymbol{x}^{n+1/2}) \\ +\boldsymbol{E}^{n+3/2}(\boldsymbol{x}^{n+1/2}) \,, \end{aligned} \tag{24}$$

where the recursive filter

$$\bar{\boldsymbol{E}}^{n+1/2} \equiv (1/2)[\bar{\boldsymbol{E}}^{n-1/2} + \boldsymbol{E}^{n+3/2}] \,. \tag{25}$$

Note that the future positions $\boldsymbol{x}^{n+3/2}$ depend on the electric field $\boldsymbol{E}^{n+3/2}$. But this field is not yet known because it is determined from the charge density $\rho^{n+3/2}$ computed from the particle positions $\boldsymbol{x}^{n+3/2}$. Thus one has a set of nonlinear coupled particle and field equations.

Techniques for making tractable the set of coupled equations in implicit schemes were developed some 15–20 years ago. They fall into two general classes: the use of moment equations and the direct approach. The motivation behind the moment approach [8], [61] is the realization that over a single time step the particle equations and the corresponding moment or fluid equations do not differ very much. One can then use an implicit set of moment equations to estimate the fields at the next time step. The particles are then advanced using these new fields, and the moment equations are reinitialized at each time step using the new particle information. The errors introduced by use of the moment equations then do not propagate. In the direct (or kinematic) approach [4], [26], [51] one predicts the future fields directly by means of linearization of the particle-field equations about an estimate (extrapolation) for their values at the new time level.

The potential advantages of the implicit approach are dramatic. Stable solutions have been obtained using time steps on the order of 50–100 times as

large as that allowed by an explicit scheme (e.g., [9], [34]). Nevertheless, because of their substantially increased complexity and concerns about their reliability, implicit schemes have not been widely used in the space physics community.

2.4 Implementation Issues

PIC simulations historically have been limited to treating configurations of limited spatial extent and temporal duration, with a relatively small number of particles per Debye cell, with only one or two spatial dimensions, and with extremely simple boundary conditions. There are a number of approaches that can overcome (at least in part) one or more of these limitations. As we have already discussed, the time step constraints can be relaxed through the use of an approximate subset of Maxwell's equations or by the choice of an implicit time integration scheme.

Uniform spatial meshes are normally employed in PIC codes and are necessary for the use of FFTs, but they are not well suited to dealing with strongly inhomogeneous systems. The chapter by *Schriver* [73] in this volume describes the use of an irregular grid in modeling a portion of the auroral zone.

While PIC models are designed to study collisionless plasmas, the models are not actually collisionless. The most important parameter in determining the numerical collision time is the number of particles per Debye cell (e.g., [35], [36]). This number is much smaller in PIC calculations, particularly in 2 and 3 dimensions, than for real plasma systems. The chapter by *Sydora* [81] in this volume describes the δf method in which particles are used to represent only the perturbed plasma distribution function. This method is particularly well suited for studying the onset of weak instabilities.

Perhaps the most significant advance in extending the capabilities of PIC models has been the development of massively parallel computers with their vastly larger effective memories. Since the dominant part ($\sim$90%) of the calculation in a PIC code involves interpolation between particles and the grid, it is essential that these two data structures reside on the same processor. The procedure used to achieve this is the general concurrent particle-in-cell (GCPIC) algorithm [55], [19]. The essential idea is that different processors are assigned different regions of the simulation grid, and particles are assigned to processors according to the spatial region they occupy. To implement this scheme with only local interpolation, it is necessary to pass some field and particle data between different processors. This is achieved either by explicit message passing (using for example MPI), or, as described in the chapter by *Cai et al.* [11] in this volume, High Performance Fortran. The development of massively parallel computing has made possible large-scale 3-D PIC calculations.

3 Hybrid Models

In a hybrid model, a full particle description is retained for either the ions or electrons, while the other species is described by fluid equations. While there

are cases where it is the electrons that are treated as particles [58], in the overwhelming majority of applications in space physics it is the ions that are the particle species. We shall consider only this case. This model is well adapted to problems involving waves in the vicinity of or below the ion cyclotron frequency. Frequently, one assumes the quasi-neutral limit where the electron charge density is equal to that of the ions and neglects the electron inertia. In all cases one neglects the displacement current, and the magnetic field is advanced using Faraday's law (3). A detailed discussion of hybrid models, including illustrative examples, is given in the chapter by *Winske et al.* [91] in this volume. Here we include a short description of the hybrid approach for completeness.

3.1 Massless Electron Models

In the simplest hybrid model, one neglects the electron mass in the generalized Ohm's law and assumes a scalar electron pressure. The electric field is then given by

$$\begin{aligned} \boldsymbol{E} = (\boldsymbol{\nabla} \times \boldsymbol{B}) \times \boldsymbol{B}/4\pi ne - \boldsymbol{u}_i \times \boldsymbol{B}/c \\ -\boldsymbol{\nabla} p_e/ne + \eta en(\boldsymbol{u}_i - \boldsymbol{u}_e) \, , \end{aligned} \tag{26}$$

where we have used $\boldsymbol{\nabla} \times \boldsymbol{B} = 4\pi \boldsymbol{J}/c$, $\boldsymbol{u}_i$ and $\boldsymbol{u}_e$ are the ion and electron fluid velocities, and η is a phenomenological resistivity describing microscopic coupling between the electrons and ions. The electron velocity $\boldsymbol{u}_e$ can be expressed in terms of $\boldsymbol{u}_i$ and $\boldsymbol{\nabla} \times \boldsymbol{B}$. To complete the fluid equations, one must specify an electron equation of state. The usual choices are isothermal or adiabatic.

The assumption of massless electrons has an important consequence for the allowable time step. Normally, the whistler wave spectrum is cutoff at the electron cyclotron frequency Ω_e. If $m_e \to 0$, however, then the whistler frequency spectrum is unbounded. This effect is most severe for parallel propagating whistlers where the frequency scales like $\omega/\Omega_i = (kc/\omega_{pi})^2$ for large k. The largest wavenumber on the grid is $k_{\max} = \pi/\Delta_g$, and the corresponding maximum phase velocity is $v_{\max} = \omega_{\max}/k_{\max} = \Omega_i k_{\max}(c/\omega_{pi})^2$. Thus the CFL constraint on the whistler waves gives

$$\Omega_i \Delta t < (\Delta_g \omega_{pi}/c)^2/\pi \, . \tag{27}$$

Since one wants to resolve the c/ω_{pi} scale, the time step is limited to a small fraction of Ω_i^{-1}.

A number of different time-advance schemes have been developed for the hybrid model (see [71] and [89] for a review). One of the early 2-D schemes is that of *Harned* [28]. Here the treatment of the particle motion employs the standard leapfrog integration, while the time advance of the field equations (3) and (26), with $\boldsymbol{E}$ and $\boldsymbol{B}$ defined at the same time interval, is accomplished using a predictor-corrector method. The electric and magnetic fields are defined on interlaced grids which have a relative separation of half a cell. To obtain spatially centered finite difference expressions, bilinear finite differences on adjacent nodes

on the interlaced grid are used. The method uses the same time step for both particles and fields and requires two passes through the ion arrays per time step.

Since the ions do not respond on the short time scale set by the whistler waves, and since the ion advance is much more time consuming than the field update, it is advantageous to use a subcycling approach in which the field equations are advanced many times (of the order of 10–50) for each advance of the ion particle variables. This subcycling approach was introduced by *Terasawa et al.* [85], who used a second-order rational Runge-Kutta algorithm for both the particles and fields. The larger time step for the ions represents a considerable computational saving and has been employed in most recent hybrid schemes (e.g., [60], [62], [80], [90]).

More recent algorithms have allowed the ions to be pushed only once per time step. *Winske and Quest* [90] describe a moment method in which the fields are subcycled using a fourth-order rational Runge-Kutta algorithm and an MHD equation is used to advance the plasma fluid velocity. This equation includes an advective term and an ion stress tensor which must be collected on the grid. *Matthews* [62] in turn avoids the need for a pre-push of the ion velocities by advancing the ion current density with an appropriate equation of motion. In this scheme the advective term and the ion stress tensor of the moment method are avoided, and the method is particularly well suited for modelling multiple ion species.

3.2 Finite Electron Inertia

In investigating the behavior of current layers at spatial scales smaller than the ion skin depth c/ω_{pi} (e.g., [23], [45], [75]), it becomes important to include the effects of the electron inertia. A modified form of the hybrid model can be derived to treat this case. A detailed discussion of such models can be found in the monograph by *Lipatov* [57]. Define the fields $\boldsymbol{B}'$ and $\boldsymbol{E}'$ by

$$\boldsymbol{B}' \equiv (1 - \delta_e^2 \nabla^2)\boldsymbol{B} \ , \tag{28}$$

$$\boldsymbol{E}' \equiv \boldsymbol{E} + (m_e/e)\partial \boldsymbol{u}_e/\partial t \ , \tag{29}$$

where $\boldsymbol{E}$ is given by the generalized Ohm's law and $\delta_e = c/\omega_{pe}$. From Ampère's Law one finds

$$(\partial/\partial t)\nabla^2 \boldsymbol{B} = (4\pi e/c)\boldsymbol{\nabla} \times (\partial/\partial t) n(\boldsymbol{u}_e - \boldsymbol{u}_i) \ . \tag{30}$$

If one neglects the unmagnetized ion response on the short spatial and temporal scales of the electron physics, that is one neglects the number density and ion current density variations, then combining (28)–(30) with Faraday's law yields

$$\partial \boldsymbol{B}'/\partial t = -c(\boldsymbol{\nabla} \times \boldsymbol{E}') \ . \tag{31}$$

This is one of the equations of what is normally described as electron magnetohydrodynamics [42].

The generalized fields $\boldsymbol{B}'$ and $\boldsymbol{E}'$ are advanced in time as in the standard hybrid model. At the end of each step, $\boldsymbol{B}$ must be unfolded from $\boldsymbol{B}'$. Here, it is usually assumed that c/ω_{pe} remains uniform both in space and time, thus neglecting any evolution in the electron scale structure [76]. The finite electron inertial correction is neglected in the electric field which is used to advance the ions forward in time [75] since this correction only becomes important at spatial scales of δ_e. Changes in the electric field on this scale have very little effect on the motion of the ions because of their large mass.

4 PIC Simulations of Collisionless Magnetic Reconnection

The conversion of energy stored in stressed magnetic fields into high speed plasma flows and thermal energy via magnetic reconnection is a fundamental process that occurs in many plasma systems possessing magnetic shear (e.g., [27], [79], [86]). In particular, magnetic reconnection plays a fundamental role in the dynamics of the magnetosphere. It facilitates the entry of particles and energy from the solar wind into the magnetosphere at the magnetopause and allows the internal magnetospheric topology to change. Magnetic reconnection relies on the presence of a dissipation mechanism in a localized region of space, the so-called diffusion region. In a sufficiently collisional plasma, resistive MHD theory is valid for describing this region and determining the reconnection rate. In the magnetosphere, however, the classical collision rate is very small. In such a collisionless plasma, the dissipation region is governed by the generalized Ohm's law, which can be written in the form

$$\boldsymbol{E} = -\frac{1}{c}\boldsymbol{u}_i \times \boldsymbol{B} + \frac{(\boldsymbol{J}\times\boldsymbol{B})/c - \boldsymbol{\nabla}\cdot\underline{\boldsymbol{P}}_e}{ne} - \frac{m_e}{e}\frac{d\boldsymbol{u}_e}{dt} \; . \tag{32}$$

Here, $\boldsymbol{u}_i(\boldsymbol{u}_e)$ is the ion (electron) fluid velocity, the current density $\boldsymbol{J} \approx en(\boldsymbol{u}_i - \boldsymbol{u}_e)$ (assuming quasineutrality), and $\underline{\boldsymbol{P}}_e$ is the electron pressure tensor. This generalized Ohm's law contains three new terms which are not present in the MHD limit: the $\boldsymbol{J}\times\boldsymbol{B}$ or Hall term, the electron pressure gradient terms, and the electron inertia terms. Each of these terms introduces new physics into the system and has an associated characteristic length scale (e.g., [22]).

The simplest magnetic field configuration for studying reconnection is the Harris neutral sheet [29]. Here the magnetic field $\boldsymbol{B} = B_x(z)\hat{x}$ reverses direction over a characteristic half thickness L,

$$B_x(z) = -B_0 \tanh(z/L) \; , \tag{33}$$

and the pressure is peaked at the null of the magnetic field so that the $\boldsymbol{J}\times\boldsymbol{B}$ force is balanced by the pressure gradient,

$$P(z) = (B_0^2/8\pi)\mathrm{sech}^2(z/L) \; . \tag{34}$$

Since the spatial variation occurs as a function of z only, this is referred to as a 1-D current sheet. In ideal MHD, where the electric field is frozen into the plasma,

the Harris 1-D current sheet is stable. In the presence of dissipation, however, the sheet becomes unstable to a tearing instability which leads to the formation of magnetic islands as a function of x along $z = 0$ [12], [53]. For a collisionless plasma, the dissipation arises from electron and/or ion Landau damping. In the Earth's magnetotail the presence of a small B_z field in addition to the main field (33) prevents the electron Landau interaction, and thus most treatments have been concerned with the ion tearing mode.

Simulation of collisionless reconnection relevant to magnetospheric configurations has been pursued for over 20 years. The representation of the reconnection physics poses a severe computational challenge. First, the simulations must be at least 2-D (x,z in magnetospheric coordinates). Second, the ion tearing mode evolves on ion spatial (the skin depth c/ω_{pi}) and temporal (the gyrofrequency Ω_i) scales. As we have discussed in Sect 2, however, with an explicit code the grid resolution and time step are set by electron physics (or the even shorter light transit time across a grid cell). Thus, the number of time steps required in the simulation will scale like m_i/m_e, and the number of grid points in each dimension will scale like $(m_i/m_e)^{1/2}$. The total cost of a 3-D simulation will thus scale like $(m_i/m_e)^{5/2}$ (in 2-D the cost scales like $(m_i/m_e)^2$). A final difficulty involves the boundary conditions. The simple 1-D current sheet (33) must be embedded into a more realistic open configuration (preferably in 3-D) in which particles and magnetic flux can escape from the local reconnection region.

In many of the early simulations of ion tearing one simply assumed that the electron dynamics were unimportant and that the electrons could thus be treated as a charge-neutralizing background (a model first introduced in [21]). This assumption is equivalent to assuming a mass ratio $m_i/m_e = 1$, so that no electrostatic effects can develop. Thus, *Terasawa* [83] studied the growth of an explosive tearing mode instability in a one component plasma. He followed the motion of 7000–10,000 particles on a 64×30 grid. A full two-species simulation was performed by *Katanuma and Kamimura* [41] using a mass ratio $m_i/m_e = 16$ with a 64×64 grid and 35,000 particles.

In the intervening years the dramatic growth in computational power has substantially increased the scope of the simulations that can be performed. In 2-D one has now achieved a rather complete understanding of the nature of collisionless reconnection and the structure of the diffusion region as a result of the GEM magnetic reconnection challenge [6], and very high resolution simulations (1600×640 grid, 2×10^8 total number of particles, mass ratio $m_i/m_e = 64$) have enabled *Hoshino et al.* [38] to determine the characteristic electron distributions in different regions of the reconnection configuration.

In the GEM magnetic reconnection challenge a standard 2-D current sheet configuration consisting of a 1-D Harris current sheet with an additional perturbation to form a neutral line was studied by a variety of simulation models: full particle, hybrid, two-fluid with and without finite electron inertia, and resistive MHD. It was found that, with the exception of resistive MHD, all the models gave nearly identical results for the reconnection rate. This rate corresponded to an inflow speed $cE_y/B_0 \approx 0.1$–$0.2\ v_A$ and was much faster than that obtained

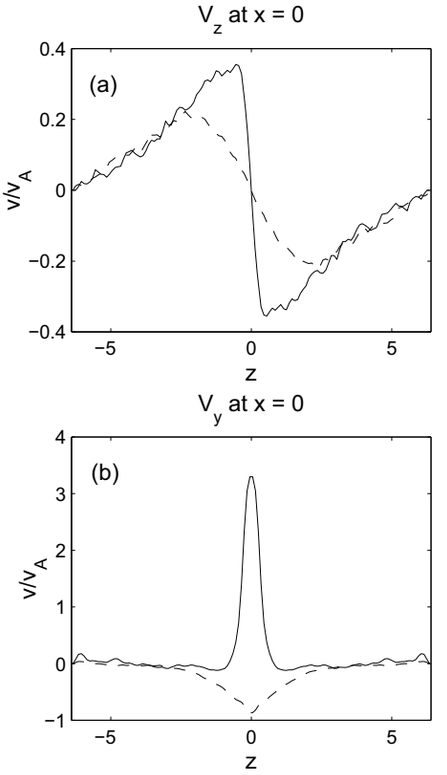


Fig. 5. Velocity profiles as a function of z at $x = 0$ averaged over all y at time $\Omega_{i0}t = 21$ for electrons (*solid lines*) and ions (*dashed lines*): (**a**) u_z, (**b**) u_y. The spatial coordinate is $z/(c/\omega_{pi}^0)$. (From [69])

in resistive MHD. The conclusion reached was that this rate is insensitive to the precise physics on the electron scale that leads to violation of the frozen-in condition (e.g., nongyrotropic electron pressure tensor, finite electron inertia, etc.), thus confirming earlier results [7], [30], [74].

In those models retaining finite electron inertia, the dissipation region develops a multiscale structure based on the ion and electron inertial lengths c/ω_{pi} and c/ω_{pe} [31], [59], [68], [75], [76]. Here we illustrate some of the features of the diffusion region based on 3-D full particle simulations [69]. For the present discussion, the y variation has been averaged out. The mass ratio is $m_i/m_e = 25$. Figure 5 shows profiles of the ion and electron flow velocities u_z and u_y as a function of z at $x = 0$. (The length scales are normalized to c/ω_{pi}^0 based on the initial peak density n_0 in the Harris distribution; the velocities are normalized to the Alfvén speed based on B_0 and n_0.) Consistent with the reconnection electric field established in the vicinity of the neutral line (see Plate 1 of [68]), the ions and electrons drift toward the neutral line from above and below. At a distance

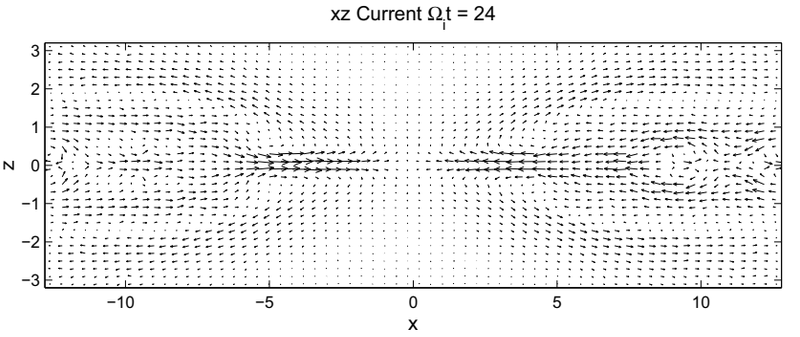


Fig. 6. Current flow vectors in the xz plane at time $\Omega_{i0}t = 24$

$|z|$ from the neutral line of about the local value of c/ω_{pi} ($\approx 2.2c/\omega^0_{pi}$) the ions become unmagnetized. The ion drift u_{iz} is then reduced in magnitude, while u_{iy} increases in magnitude in response to the direct E_y field. The electrons remain magnetized down to the much smaller distance of the local c/ω_{pe} ($\approx 2.5c/\omega^0_{pe}$) where they also respond to the direct E_y. The electron drift u_{ey} reaches $\approx 3.5v_A$. Thus the electrons are the dominant carrier of the cross-field current in a narrow region around the neutral line.

As the electrons and ions approach the neutral line, they are diverted and accelerated outward by a complex parallel electric field structure (see Plate 1 of [69]). The resulting electron and ion flow velocities are unequal, thus producing a characteristic pattern of in-plane or Hall currents (Fig. 6). The electrons are expelled from the neutral line in a narrow region of half width $\sim c/\omega_{pe}$ at speeds several times v_A. This constitutes a current in toward the neutral line for $|x| \lesssim 5c/\omega^0_{pi}$. This central current is fed at $|x| \gtrsim 5c/\omega^0_{pi}$ by an outward electron flow just inside the magnetic separatrix. The current away from the neutral line is carried by electrons flowing inward outside the separatrix. The Hall currents produce the characteristic quadrupolar "out-of-plane" B_y magnetic field [79], [84] (see Plate 2 of [69]), which has a peak magnitude of about 0.3 B_0.

With the advent of massively parallel computers in the past few years, it has become possible to extend the full particle reconnection simulations to 3-D. An issue of current research is to determine whether the laminar current and density structures observed in the 2-D solutions are preserved in 3-D or whether they become turbulent and break up under the influence of finite-k_y modes. Using two-fluid simulations, *Rogers et al.* [72] found that the structures did break up due to the electron shear flow and lower hybrid drift instabilities. Here we examine this question using a 3-D full particle simulation. In the present runs the appropriate symmetry/antisymmetry conditions have been imposed at $z = 0$ in order to prevent excitation of the drift kink mode [69], since it has been determined that the linear growth rate for this mode should be extremely small for a realistic value of the mass ratio m_i/m_e [15]. (Note, however, that recent work [52], [16] has demonstrated that the nonlinear evolution of the lower hybrid

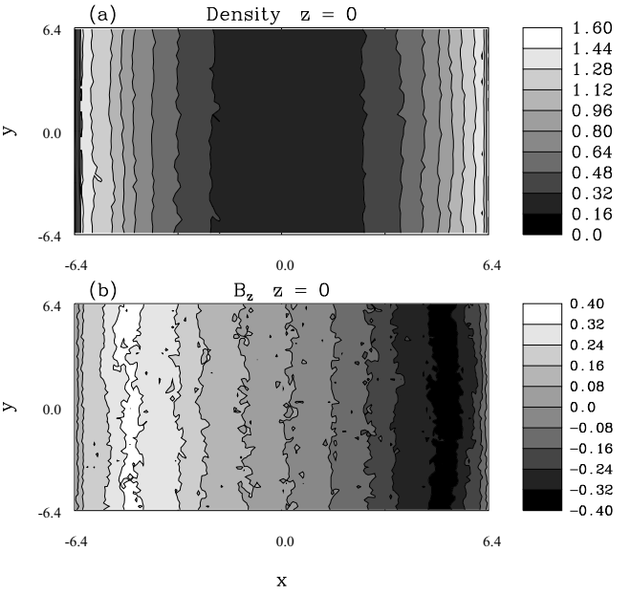


Fig. 7. Contours in the xy plane at $z = 0$ at time $\Omega_{i0}t = 17.5$ of (**a**) the density and (**b**) the magnetic field component B_z

drift instability can lead to the growth of kink modes that is more rapid than predicted by linear theory when $m_i/m_e > 100$.) These symmetry conditions at $z = 0$ would allow the growth of sausage type modes, but no such modes are observed in the simulation. This is consistent with a recent analysis [78] that showed that the electrostatic interaction is strongly stabilizing for this type of mode. The x boundaries are "open" [69]: particles crossing such a boundary are removed from the system, new thermal distributions of particles are injected, and it is required that the perturbed magnetic field $\delta B_z \equiv 0$. This latter condition allows magnetic flux to cross the boundary, but the simulation must be stopped once the tearing magnetic field perturbation reaches the boundary. The system dimensions are $-6.4c/\omega_{pi}^0 \le x \le 6.4c/\omega_{pi}^0$, $-3.2c/\omega_{pi}^0 \le y, z \le 3.2c/\omega_{pi}^0$, the mass ratio is $m_i/m_e = 100$, and the temperature ratio is $T_i/T_e = 5$.

Figure 7 shows contour plots of the density and magnetic field in the $z = 0$ plane at time $\Omega_{i0}t = 17.5$. The development of the neutral line at $x \approx 0$ is quite clear; the density in this region is strongly reduced and there is no indication of any significant structure in the y direction. This result is confirmed in Fig. 8 which shows an isosurface plot of $B_x/B_0 = -0.1$ at the same time. In the initial Harris equilibrium this surface is located at $z = 0.05c/\omega_{pi}^0$. The distance of this surface from the $z = 0$ midplane is inversely proportional to $|\partial B_x/\partial z|$ and is thus a measure of J_y. (Near $z = 0$, $|\partial B_x/\partial z| \gg |\partial B_z/\partial x|$.) Near $z = 0$, the value of $|\partial B_x/\partial z|$ has been reduced to about 75% of its initial value. This current reduction is associated with the sharply reduced density ($\approx 0.26n_0$)

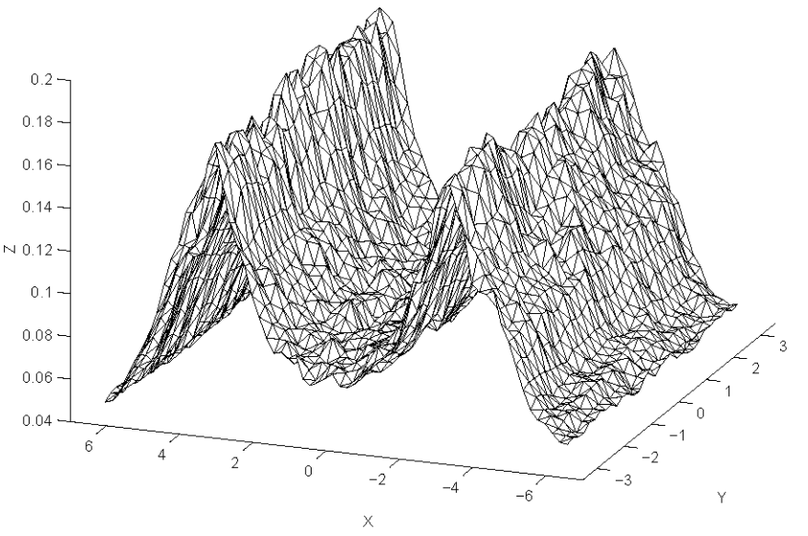


Fig. 8. Isosurface plot in 3D for $B_x/B_0 = -0.1$ at time $\Omega_{i0}t = 17.5$. Note that the scale on the x axis runs from right to left

present as lobe plasma reconnects, which more than offsets the large electron u_y drift velocity ($\approx 5v_A$) near $z = 0$. The value of $|\partial B_x/\partial z|$ continues to decrease out to $|x| \sim 3$–$4\ c/\omega_{pi}^0$, which represents the edge of the diffusion region. Here the inductive E_y field reaches its largest magnitude ($(c/v_A)E_y/B_0 \approx 0.25$). There is then a rapid increase in $|\partial B_x/\partial z|$ (decrease in distance from $z = 0$) over a distance of 1–2 c/ω_{pi}^0. This change is associated with the sharp increase in density seen in Fig. 7. Overall there is virtually no y variation in the isosurface plot; thus the reconnection fields in the diffusion region remain basically two-dimensional.

In the future one needs to extend the 3-D PIC simulations of magnetic reconnection to values of m_i/m_e closer to the proton/electron value of 1836 (simulations over a reduced 3-D volume using $m_i/m_e = 278$ have been reported by *Drake et al.* [24] and *Zeiler et al.* [93]) and to larger spatial regions. There is some evidence [65], [69] that the reconnection outflows can become unstable to an ideal pressure-gradient instability with the character of a kink or interchange mode with a wavelength satisfying $k_y\rho_i \sim 0.3$–0.4. This may lead to the formation of discrete channels of flow away from the neutral line. In addition, it may be possible that kink and/or Kelvin-Helmholtz type modes can affect nonlinearly the reconnection process for values of $m_i/m_e > 100$ [77].

5 Discussion

Both the successes and limitations of the particle-in-cell approach to plasma simulation are related to its method of mimicing real plasma behavior. With a PIC model, one is truly simulating a plasma: as a collection of particles immersed

in and evolving with an electromagnetic field that they self-consistently produce. The time evolution of the PIC model is achieved through a large number of steps in each of which all particles and field components are advanced a finite step. The PIC model reproduces all of the complex collective and nonlinear phenomena of a collisionless plasma, and, unlike the solutions of fluid equations, the density of particles can never become negative.

Traditionally, particle-in-cell simulations have suffered from the defects of treating highly idealized configurations with only one or two spatial dimensions (thus implying that the particles are actually planar sheets or line elements and that the spatial dependence of the electromagnetic fields is modified accordingly) and extremely simple boundary conditions such as periodicity (thus implying that the modeled region is a small part of an infinitely repeated system). This is beginning to change. 2-D PIC simulations are now standard, and a growing number of 3-D simulations have been performed. This has been made possible by the development of massively parallel computers and algorithms for parallel PIC simulations [55]. The current generation of parallel codes typically employ on the order of 100 processors and use a 1-D domain decomposition. Since the next few years will bring several orders of magnitude increase in the number of available processors, one can anticipate a tremendous increase in the size and complexity of the problems that can be addressed with PIC simulations. This will clearly require the development of efficient multi-dimensional domain decompositions which are compatible with the desired boundary conditions and the requisite field solutions.

Acknowledgments

The preparation of this article was supported by NASA grants NAG 5-4339, NAG 5-8060, and NAG 5-10287. The particle simulations were performed at the San Diego Supercomputer Center, which is supported by the National Science Foundation, and at the National Energy Research Scientific Computing Center at the Lawrence Berkeley National Laboratory.

References

1. B.S. Alder, S. Fernbach, M. Rotenberg (Eds.): *Methods in Computational Physics*, vol. 9 (Academic, New York 1970)
2. B.S. Alder, S. Fernbach, M. Rotenberg (Eds.): *Methods in Computational Physics*, vol. 16 (Academic, New York 1976)
3. M. Ashour-Abdalla, D.A. Dutton (Eds.): *Space Plasma Simulations* (D. Reidel, Dordrecht 1985)
4. D.C. Barnes, T. Kamimura, J.-N. Leboeuf, T. Tajima: J. Comput. Phys. **52**, 480 (1983)
5. C.K. Birdsall, A.B. Langdon: *Plasma Physics Via Computer Simulation* (McGraw-Hill, New York 1985)
6. 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, P.L. Pritchett: J. Geophys. Res. **106**, 3715 (2001)

7. D. Biskamp, E. Schwarz, J.F. Drake: Phys. Plasmas **4**, 1002 (1997)
8. J.U. Brackbill, D.W. Forslund: J. Comput. Phys. **46**, 271 (1982)
9. J.U. Brackbill, H.X. Vu: Geophys. Res. Lett. **20**, 2015 (1993)
10. J. Busnardo-Neto, P.L. Pritchett, A.T. Lin, J.M. Dawson: J. Comput. Phys. **23**, 300 (1977)
11. D. Cai, Y. Li, K-I. Nishikawa, C. Xiao, X. Yan, and Z. Pu: 'Parallel TRISTAN Code Using High Performance Fortran'. In: *Space Plasma Simulations*, this volume (Springer-Verlag, Heidelberg 2002)
12. B. Coppi, G. Laval, R. Pellat: Phys. Rev. Lett. **16**, 1207 (1966)
13. R. Courant, K.O. Friedrichs, H. Lewy: Math. Ann. **100**, 32 (1928)
14. C.G. Darwin: Philos. Mag. **39**, 537 (1920)
15. W. Daughton: Phys. Plasmas **6**, 1329 (1999)
16. W. Daughton: Phys. Plasmas **9**, in press (2002)
17. J.M. Dawson: Phys. Fluids **5**, 445 (1962)
18. J.M. Dawson: Rev. Mod. Phys. **55**, 403 (1983)
19. V.K. Decyk: Comput. Phys. Commun. **87**, 87 (1995)
20. V.K. Decyk, J.M. Dawson: J. Comput. Phys. **30**, 407 (1979)
21. D.O. Dickman, R.L. Morse, C.W. Nielson: Phys. Fluids **12**, 1708 (1969)
22. J.F. Drake: 'Magnetic Reconnection: A Kinetic Treatment'. In: *Physics of the Magnetopause*, ed. by P. Song, B.U.Ö. Sonnerup, M.F. Thomsen (AGU, Washington, D.C., 1995) pp. 155-165
23. J.F. Drake, R.G. Kleva, M.E. Mandt: Phys. Rev. Lett. **73**, 1251 (1994)
24. J.F. Drake, B.N. Rogers, M.A. Shay: 'The Physics of Collisionless Magnetic Reconnection'. Presented at: The First S-RAMP Conference, Sapporo, Japan, p. 317 (2000)
25. D.W. Forslund: Space Sci. Rev. **42**, 3 (1985)
26. A. Friedman, A.B. Langdon, B.I. Cohen: Comments Plasma Phys. Controlled Fusion **6**, 225 (1981)
27. A.A. Galeev: 'Spontaneous Reconnection of Magnetic Field Lines in a Collisionless Plasma'. In: *Basic Plasma Physics*, vol. 2, ed. by A.A. Galeev, R.N. Sudan (North-Holland, Amsterdam 1984), pp. 305–335
28. D.S. Harned: J. Comput. Phys. **47**, 452 (1982)
29. E.G. Harris: Nouvo Cimento **23**, 115 (1962)
30. M. Hesse, K. Schindler, J. Birn, M. Kuznetsova: Phys. Plasmas **6**, 1781 (1999)
31. M. Hesse, J. Birn, M. Kuznetsova: J. Geophys. Res. **106**, 3721 (2001)
32. D.W. Hewett: Space Sci. Rev. **42**, 29 (1985)
33. D.W. Hewett, J.B. Boyd: J. Comput. Phys. **70**, 166 (1987)
34. D.W. Hewett, G.E. Frances, C.E. Max: Phys. Rev. Lett. **61**, 893 (1988)
35. R.W. Hockney: J. Comput. Phys. **8**, 19 (1971)
36. R.W. Hockney, S.P. Goel, J.W. Eastwood: J. Comput. Phys. **14**, 148 (1974)
37. R.W. Hockney, J.W. Eastwood: *Computer Simulation Using Particles* (McGraw-Hill, New York 1981)
38. M. Hoshino, K. Hiraide, T. Mukai: Earth, Planets, Space **53**, 627 (2001)
39. J.D. Jackson: *Classical Electrodynamics*, 2nd Ed. (Wiley, New York 1975) pp. 593-595
40. G. Joyce, M. Lampe, G. Ganguli: 'Particle Simulation of Dusty Plasmas'. In: *Space Plasma Simulations*, this volume (Springer-Verlag, Heidelberg 2002)
41. I. Katanuma, T. Kamimura: Phys. Fluids **23**, 2500 (1980)
42. A.S. Kingsep, K.V. Chukbar, V.V. Yan'kow: 'Electron Magnetohydrodynamics'. In: *Review of Plasma Physics*, Vol. 16, ed. by B.B. Kadomtsev (Consultants Bureau, New York 1990) pp. 243–291

43. A.J. Klimas: J. Comput. Phys. **68**, 202 (1987)
44. A.J. Klimas, W.M. Farrell: J. Comput. Phys. **110**, 150 (1994)
45. M.M. Kuznetsova, M. Hesse, D. Winske: J. Geophys. Res. **103**, 199 (1998)
46. A.B. Langdon: J. Comput. Phys. **6**, 247 (1970)
47. A.B. Langdon: J. Comput. Phys. **30**, 202 (1979)
48. A.B. Langdon: Comput. Phys. Commun. **70**, 447 (1992)
49. A.B. Langdon, D.C. Barnes: 'Direct Implicit Plasma Simulation'. In: *Multiple Time Scales*, ed. by J.U. Brackbill, B.I. Cohen (Academic, Orlando, FL 1985) pp. 335-375
50. A.B. Langdon, B.F. Lasinski: 'Electromagnetic and Relativistic Plasma Simulation Models'. In: *Methods in Computational Physics*, Vol. 16, ed. by B. Alder, S. Fernbach, M. Rotenberg (Academic, New York 1976) pp. 327–366
51. A.B. Langdon, B.I. Cohen, A. Friedman: J. Comput. Phys. **51**, 107 (1983)
52. G. Lapenta, J.U. Brackbill: Phys. Plasmas **9**, 1544 (2002)
53. G. Laval, R. Pellat, M. Vuillemin: 'Instabilités Electromagnétiques des Plasmas Sans Collisions'. In: *Plasma Physics and Controlled Fusion Research*, vol. II, (International Atomic Energy Agency, Vienna 1966) pp. 259–277
54. B. Lembège, J.W. Eastwood (Eds.): *Tutorial Courses: Third International School for Space Simulation* (Cepadues, Toulouse 1989)
55. P.C. Liewer, V.K. Decyk: J. Comput. Phys. **85**, 302 (1989)
56. E.L. Lindman: J. Comput. Phys. **5**, 13 (1970)
57. A.S. Lipatov: *The Hybrid Multiscale Simulation Technology* (Springer-Verlag, Heidelberg 2001)
58. P.M. Lyster, J.-N. Leboeuf: J. Comput. Phys. **102**, 180 (1992)
59. Z.W. Ma, A. Bhattacharjee: J. Geophys. Res. **106**, 3773 (2001)
60. M.E. Mandt, R.E. Denton, J.F. Drake: Geophys. Res. Lett. **21**, 73 (1994)
61. R.J. Mason: J. Comput. Phys. **41**, 233 (1981)
62. A.P. Matthews: J. Comput. Phys. **112**, 102 (1994)
63. H. Matsumoto, Y. Omura (Eds.): *Computer Space Plasma Physics: Simulation Techniques and Software* (Terra, Tokyo 1993)
64. H. Matsumoto, T. Sato (Eds.): *Computer Simulations of Space Plasmas* (Reidel, Dordrecht 1984)
65. M.S. Nakamura, M. Fujimoto, H. Matsumoto: Adv. Space Res., in press (2001)
66. C.W. Nielson, H.R. Lewis: 'Particle Models in the Nonradiative Limit'. In: *Methods in Computational Physics*, Vol. 16, ed. by B. Alder, S. Fernbach, M. Rotenberg (Academic, New York 1976) pp. 367–388
67. M. Ozaki, T. Sato, R. Horiuchi, the Complexity Simulation Group: Phys. Plasmas **3**, 2265 (1996)
68. P.L. Pritchett: J. Geophys. Res. **106**, 3783 (2001)
69. P.L. Pritchett: J. Geophys. Res. **106**, 25,961 (2001)
70. P.L. Pritchett, F.V. Coroniti, V.K. Decyk: J. Geophys. Res. **101**, 27,413 (1996)
71. K.B. Quest: 'Hybrid Simulation'. In: *Tutorial Courses: Third International School for Space Simulation*, ed. by B. Lembège, J.W. Eastwood (Cepadues, Toulouse 1989) p. 177
72. B.N. Rogers, J.F. Drake, M.A. Shay: Geophys. Res. Lett. **27**, 3157 (2000)
73. D. Schriver: 'Simulating an Inhomogeneous Plasma System: Variable Grids and Boundary Conditions'. In: *Space Plasma Simulations*, this volume (Springer-Verlag, Heidelberg 2002)
74. M.A. Shay, J.F. Drake: Geophys. Res. Lett. **25**, 3759 (1998)
75. M.A. Shay, J.F. Drake, R.E. Denton, D. Biskamp: J. Geophys. Res. **103**, 9165 (1998)

76. M.A. Shay, J.F. Drake, B.N. Rogers, R.E. Denton: J. Geophys. Res. **106**, 3759 (2001)
77. I. Shinohara, H. Suzuki, M. Fujimoto, M. Hoshino: Phys. Rev. Lett. **87**, 095001 (2001)
78. I. Silin, J. Büchner, L. Zelenyi: Phys. Plasmas **9**, 1104 (2002)
79. B.U.Ö. Sonnerup: 'Magnetic Field Reconnection'. In: *Solar System Plasma Physics*, vol. III, ed. by L.T. Lanzerotti, C.F. Kennel, E.N. Parker (North-Holland, Amsterdam 1979) pp. 45–108
80. D.W. Swift: J. Comput. Phys. *126*, 109 (1996)
81. R.D. Sydora: 'Low Noise Electrostatic and Electromagnetic δF Particle-in-Cell Plasma Simulation Method'. In: *Space Plasma Simulations*, this volume (Springer-Verlag, Heidelberg 2002)
82. T. Tajima: *Computational Plasma Physics: With Applications to Fusion and Astrophysics* (Addison-Wesley, Redwood City, CA, 1989)
83. T. Terasawa: J. Geophys. Res. **86**, 9007 (1981)
84. T. Terasawa: Geophys. Res. Lett. **10**, 475 (1983)
85. T. Terasawa, M. Hoshino, J.-I. Sakai, T. Hada: J. Geophys. Res. **91**, 4171, (1986)
86. V.M. Vasyliunas: Rev. Geophys. **13**, 303 (1975)
87. J. Villasenor, O. Buneman: Comput. Phys. Commun. **69**, 447 (1992)
88. J. Wang, P. Liewer, V. Decyk: Comput. Phys. Commun. **87**, 35 (1995)
89. D. Winske, N. Omidi: 'Hybrid Codes: Methods and Applications'. In: *Computer Space Plasma Physics: Simulation Techniques and Software*, ed. by H. Matsumoto, Y. Omura (Terra, Tokyo 1993) pp. 103–160
90. D. Winske, K.B. Quest: J. Geophys. Res. **93**, 9681 (1988)
91. D. Winske, L. Yin, N. Omidi: 'Hybrid Codes: Past, Present and Future'. In: *Space Plasma Simulations*, this volume (Springer-Verlag, Heidelberg 2002)
92. K.S. Yee: IEEE Trans. Antennas Propagat. **14**, 302 (1966)
93. A. Zeiler, D. Biskamp, J.F. Drake, B.N. Rogers, M. Shay, M. Scholer: 'Multigrid Parallel PIC Coding: Method and Results'. In: *Space Plasma Simulation: Proc. Sixth International School/Symposium ISSS-6*, edited by J. Büchner, C.T. Dum, M. Scholer, (Copernicus Gesellschaft, Katlenburg-Lindau 2001) pp. 8–11
94. Z. Zhu, R.M. Winglee: J. Geophys. Res. **101**, 4885 (1996)

Parallel 3-D Electromagnetic Particle Code Using High Performance FORTRAN: Parallel TRISTAN

Dongsheng Cai[1], Yaoting Li[1,3], Ken-Ichi Nishikawa[2], Chijie Xiao[1,3], Xiaoyang Yan[1], and Zuying Pu[3]

[1] Institute of Information Sciences and Electronics,
The University of Tsukuba, Ibaraki 305-8573, Japan
email:{cai, ytli, cjxiao, yxy }@is.tsukuba.ac.jp
[2] National Space Science and Technology Center, 320 Sparkman Drive, SD 50,
Huntsville, AL 35805 USA email: Ken-Ichi.Nishikawa@nsstc.nasa.gov
[3] Department of geophysics, Peking University,
Beijing 100871, China
email:zypu@pku.edu.cn

Abstract. A three-dimensional full electromagnetic particle-in-cell (PIC) code, TRISTAN (Tridimensional Stanford) code, has been parallelized using High Performance Fortran (HPF) as a RPM (Real Parallel Machine). In the parallelized HPF code, the simulation domain is decomposed in one-dimension, and both the particle and field data located in each domain that we call the sub-domain are distributed on each processor. Both the particle and field data on a sub-domain are needed by the neighbor sub-domains and thus communications between the sub-domains are inevitable. Our simulation results using HPF exhibit the promising applicability of the HPF communications to a large scale scientific computing such as solar wind-magnetosphere interactions.

1 Introduction

This paper reports on parallelization of Tridimensional Stanford (TRISTAN) code (Buneman, 1993) that is a three-dimensional electromagnetic full particle code developed at Stanford University on a two-way PentiumPro PC cluster that consists of 16 distributed SMPs and other commercial parallel computers like Fujitsu VPP5000, NEC SX-6 and Hitachi SR-8000 etc. using High Performance Fortran (HPF).

In our parallel program, the simulation domain is decomposed into the sub-domains as shown in Fig. 1. The Particle-In-Cell (PIC) computation in TRISTAN to be performed on a certain sub-domain or on a certain processor where the sub-domain is distributed will typically require the data from their neighbor processors to proceed the whole PIC simulations. Here we distribute the field arrays and the particles over processors as indicated in Fig. 1. Thus the data must be transferred between processors in each time step so as to allow PIC simulation to proceed in time. These inter-processor communications in each time step need to be programmed in HPF constructs.

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 25–53, 2003.

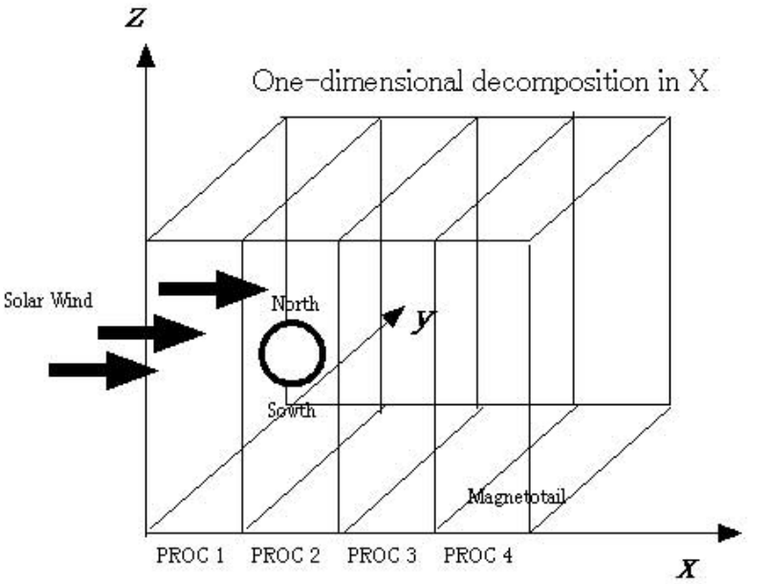


Fig. 1. Coordinate of the simulation domains and domain decomposition in x.

The amount of inter-processor communications needed for a parallel program basically depends on the algorithms and the scales of the physical problem sizes adopted in the simulations. In PIC simulations, they are the way decomposing the simulation domains, the sizes of the sub-domain boundaries, and the number of the particles in a cell, respectively.

The pgHPF compiler of Portland Group Inc. aims to realize the standard High Performance Fortran specification and can be installed on a number of parallel machines. Executable codes produced by the pgHPF compilers for PCs with IA-32 CPUs are unconstrained, and can be executed on any compatible IA-32 processor-based system regardless of whether the pgHPF compilers are installed on that system or not. From the HPF programmer's point of view, the differences between versions of the pgHPF runtime library have little effect on program developments.

In parallel programming models, usually, the SPMD (Single Program Multiple Data) models using MPI (Message Passing Interface) or PVM (Parallel Virtual Machine) are one of the most popular models. Our HPF TRISTAN code also uses the same SPMD models. The biggest advantage of HPF is its programming style. Once the simulation domain is decomposed and the data are distributed to each sub-domains or over processors using simple HPF compiler directives, other HPF programming styles are very similar to those in usual Fortrans. Of course, the biggest problem here is the performance issues comparing with those using MPI or PVM.

Actually, pgHPF is based on a RPM (PGI Proprietary Communications - Real Parallel Machine) protocol. This transport mechanism was developed by PGI to model the behavior of PVM among a homogeneous group of hosts on a network. It offers both greater programming efficiency and performance than PVM with fewer requirements. In this paper, to archive a similar high performance using HPF comparing with that using MPI or PVM in the full electromagnetic PIC simulation, some careful optimizations of inter-processor communications are proposed.

Our code is the same as TRISTAN code except for the parallelization part, which utilizes rigorous charge-conserving formulas and radiating boundary conditions (Buneman, 1993). It was written in HPF so that the code can be run on any parallel computers with the HPF compilers.

The parallelization part of our HPF TRISTAN code is same as (Liewer and Decyk, 1985) and (Decyk, 1995). We separate the communication parts from computation parts, and use both the "particle manager" and the "field manager" to localize the inter-processor communications (Decyk, 1995) as shown in Fig. 2. Thus the code can be easily converted to MPI or PVM version of TRISTAN code.

The basic controlling equations of the plasmas are Newton-Lorentz equation:

$$m_{i,e}\frac{d\boldsymbol{v}_{i,e}}{dt} = q_{i,e}(\boldsymbol{E} + \boldsymbol{v}_{i,e} \times \boldsymbol{B}), \tag{1}$$

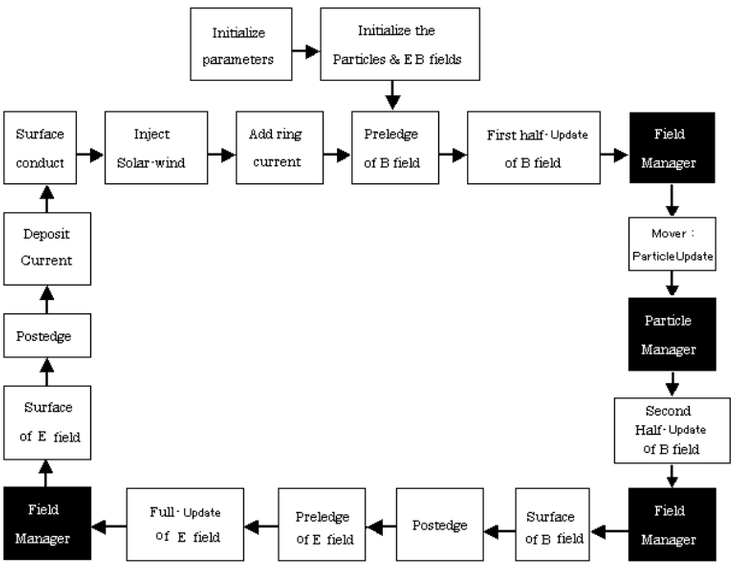


Fig. 2. The computational cycle of the HPF TRISTAN code. The black boxes represent HPF communications.

where i and e corresponds to ion and electron, respectively, and Maxwell equations:

$$\frac{\partial \boldsymbol{B}}{\partial t} = - \bigtriangledown \times \boldsymbol{E}, \tag{2}$$

$$\frac{\partial \boldsymbol{E}}{\partial t} = c^2 \bigtriangledown \times \boldsymbol{B} - \frac{1}{\varepsilon_0} \boldsymbol{J}, \tag{3}$$

Here

$$\boldsymbol{J} = \sum (n_i q_i \boldsymbol{v}_i - n_e q_e \boldsymbol{v}_e). \tag{4}$$

The coordinate and one-dimensional domain decomposition using in the simulation domain is shown in Fig. 1. For parallel benchmarking purposes, we perform the real simulations of solar wind-magnetosphere interactions using the code. For the simulation of solar wind-magnetosphere interactions, the following boundary conditions were used for the particles (Buneman, 1993): (1) Fresh particles representing the incoming solar wind (unmagnetized in our test run) are continuously injected across the yz plane at $x = x_{min}$ with a thermal velocity plus a bulk velocity in the $+x$ direction; (2) Thermal solar particle flux is also injected across the sides of our rectangular computation domain; (3) Escaping particles are arrested in a buffer zone, redistributed in those grid they escaped more uniformly by making the zone conducting in order to simulate their escape to infinity, and finally written off. We use a simple model for the ionosphere where both electrons and ions are reflected by the Earth dipole magnetic field. The effects of the Earth rotation are not included. The effect of thermal expansion of the solar-wind is also not included. Since the solar-wind and the Earth dipole magnetic field are included, some load-imbalance due to this asymmetry is expected in this HPF TRISTAN code. In Sect. 2, some basics of TRISTAN code and the ways to run it are introduced. In Sect. 3, basics of HPF TRISTAN data structure and array distributions are discussed. In Sects. 4 and 5, field and particle data domain decompositions and the way of communication between processors are discussed, respectively. In Sect. 6, unstability of the HPF communication and the way that avoids the performance degradation are discussed. In Sect. 7, benchmark and simulation results of the HPF TRISTAN code on PC cluster are discussed. Section 8 concludes the remarks of this paper.

2 Basics of TRISTAN Code

The control equations of TRISTAN code are Maxwell and Newton-Lorentz equations only. Instead of solving Poisson equation that is solved numerically in almost all particle simulation codes, TRISTAN code solves only two curls, i.e. Ampere and Faraday equations. A rigorous charge conservation method for the current deposits is described in (Villasenor and Buneman, 1992).

The particles that are initialized as unmagnetized Maxwell distribution are updated by the leap-frog method. Throughout the code the linear interpolation is employed. The computational cycle of HPF TRISTAN code is displayed in Fig. 2. The black boxes in the figure are HPF communication subroutines and they are field and particle managers.

2.1 Fields

TRISTAN code scales such that $\epsilon_0 = 1$ and hence $\mu_0 = 1/c^2$. This also means $\mathbf{E} = \mathbf{D}$. Instead of recording components of $\mathbf{B}$ or $\mathbf{H}$, TRISTAN records bx, by, bz of $c\mathbf{B}$ (alias $\mathbf{H}/\mathbf{c}$). This makes symmetry for electric field and magnetic field ($\mathbf{E} \longleftrightarrow \mathbf{B}$) in Maxwell equations. Throughout, TRISTAN uses a rectangular cubic grid with $\delta x = \delta y = \delta z = 1$ and time discretisation with $\delta t = 1$. Before and after moving (or pushing) the particles, $\mathbf{B}$ is updated in two half steps so that it is available at the same time as $\mathbf{E}$ for the particle update.

In TRISTAN code, only two curls of Maxwell equations are solved:

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E}, \tag{5}$$

$$\frac{\partial \mathbf{D}}{\partial t} = \nabla \times \mathbf{H} - \mathbf{J}. \tag{6}$$

and here

$$\mathbf{B} = \mu_0 \mathbf{H}, \tag{7}$$

$$\mathbf{D} = \epsilon_0 \mathbf{E}. \tag{8}$$

If we scale Maxwell equations using $\epsilon_0 = 1$, and substitute $\mathbf{E} = \mathbf{D}$, $\mathbf{B} = \mathbf{cB}$ into eq. (5) and (6), we obtain:

$$\frac{\partial}{\partial t}\mathbf{B} = -c\nabla \times \mathbf{E}, \tag{9}$$

$$\frac{\partial}{\partial t}\mathbf{E} = c\nabla \times \mathbf{B} - \mathbf{J}. \tag{10}$$

These two equations imply the fields symmetry($\mathbf{E} \longleftrightarrow \mathbf{B}$).

2.2 Magnetic Field Update

The staggered grid mesh system, known in the computational electromagnetic community as Yee lattice (Yee, 1966), is shown in Fig. 3. It ensures that the change of $\mathbf{B}$ flux through a cell surface equals the negative circulation of $\mathbf{E}$ around that surface and the change of $\mathbf{E}$ flux through a cell surface equals the circulation of $\mathbf{B}$ around that surface minus the current through it. Here $\mathbf{B}$ and $\mathbf{E}$ are in a symmetry form except subtracting the charge flux $\mathbf{J}$ in Ampere equation. Charge flux $\mathbf{J}$ is calculated and subtracted after the particles are moved later in the program. Thus magnetic fields are updated as follows:

The change of $\mathbf{B}$ flux can be expressed as:

$$\frac{\partial \mathbf{B}}{\partial t} = -c \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ e_x & e_y & e_z \end{vmatrix} = c[\ \mathbf{i}(\frac{\partial e_y}{\partial z} - \frac{\partial e_z}{\partial y}) + \mathbf{j}(\frac{\partial e_z}{\partial x} - \frac{\partial e_x}{\partial z}) + \mathbf{k}(\frac{\partial e_x}{\partial y} - \frac{\partial e_y}{\partial x})] \tag{11}$$

In Yee lattice, e_x, e_y, e_z, b_x, b_y, and b_z are, respectively, staggered and shifted on 0.5 from (i, j, k) and located at the positions as follows:

$$\begin{array}{l} e_x(i,j,k) \rightarrow e_x(i+.5,j,k), \\ e_y(i,j,k) \rightarrow e_y(i,j+.5,k), \\ e_z(i,j,k) \rightarrow e_z(i,j,k+.5), \end{array} \tag{12}$$

and

$$\begin{array}{l} b_x(i,j,k) \rightarrow b_x(i,j+.5,k+.5), \\ b_y(i,j,k) \rightarrow b_y(i+.5,j,k+.5), \\ b_z(i,j,k) \rightarrow b_z(i+.5,j+.5,k). \end{array} \tag{13}$$

In our simulation, we use integer grids. In both Eq. (12) and (13), i, j, k in the right-hand sides correspond to Fortran array indices notations and i, j, k in the left hand sides correspond to the real positions in the simulation domains as shown in Fig. 3. In this report, if the values "0.5" are added to either i, j, k in the array indices, then the array indices correspond to the real positions in the simulation domains.

Thus the magnetic field components b_x, b_y, b_z are, respectively, updated by the negative circulation of $\mathbf{E}$ around Yee lattice surface as follows:

$$\begin{array}{l} \frac{\partial}{\partial t} b_x = (b_x^{new}(i,j+.5,k+.5) - b_x^{old}(i,j+.5,k+.5))/\delta t \\ \quad = c[(e_y(i,j+.5,k+1) - e_y(i,j+.5,k))/\delta z \\ \quad\quad -(e_z(i,j+1,k+.5) - e_z(i,j,k+.5))/\delta y]. \end{array} \tag{14}$$

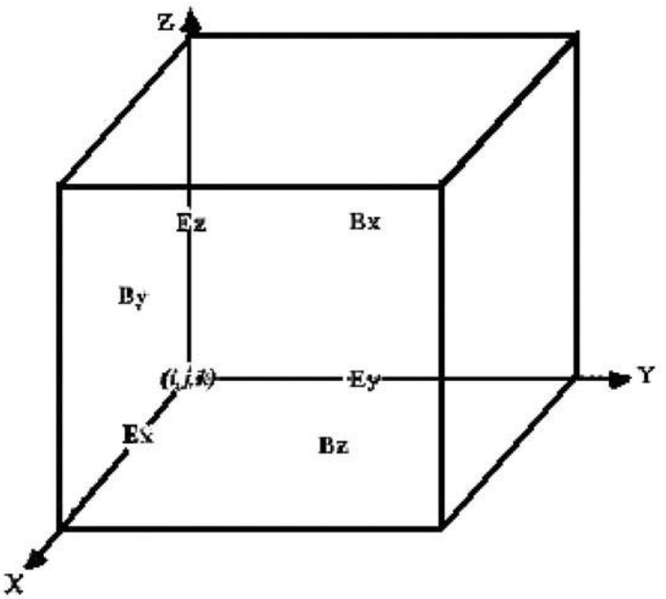


Fig. 3. The positions of field components in Yee lattice.

Here $\delta t = \delta z = \delta y = \delta x = 1$. Thus we get the update form:

$$\begin{aligned} b_x^{new}(i,j,k) &= b_x^{old}(i,j,k) \\ &+c[\ e_y(i,j,k+1) - e_y(i,j,k) - e_z(i,j+1,k) + e_z(i,j,k)\ ]. \end{aligned} \tag{15}$$

To get the update form of b_y, and b_z, the same procedures are as followed:

$$\begin{aligned} \tfrac{\partial}{\partial t} b_y &= (b_y^{new}(i+.5,j,k+.5) - b_y^{old}(i+.5,j,k+.5))/\delta t \\ &= c[(e_z(i+1,j,k+.5) - e_z(i,j,k+.5))/\delta x \\ &- (e_x(i+.5,j,k+1) - e_x(i+.5,j,k))/\delta z\ ], \end{aligned} \tag{16}$$

$$\begin{aligned} b_y^{new}(i,j,k) &= b_y^{old}(i,j,k) \\ &+c[\ e_z(i+1,j,k) - e_z(i,j,k) - e_x(i,j,k+1) + e_x(i,j,k)\ ], \end{aligned} \tag{17}$$

$$\begin{aligned} \tfrac{\partial}{\partial t} b_z &= (b_z^{new}(i+.5,j+.5,k) - b_z^{old}(i+.5,j+.5,k))/\delta t \\ &= c[(e_x(i+.5,j+1,k) - e_x(i+.5,j,k))/\delta y \\ &- (e_y(i+1,j+.5,k) - e_y(i,j+.5,k))/\delta x], \end{aligned} \tag{18}$$

$$\begin{aligned} b_z^{new}(i,j,k) &= b_z^{old}(i,j,k) \\ &+c[\ e_x(i,j+1,k) - e_x(i,j,k) - e_y(i+1,j,k) + e_y(i,j,k)]. \end{aligned} \tag{19}$$

2.3 Electric Field Update

In Yee lattice, e_x, e_y, and e_z are, respectively, staggered and shifted 0.5 from (i,j,k) and located at the positions as shown in Fig. 3.

The change of **E** flux through a cell surface equals the circulation of **B** around that surface minus the current through it. First, the electric field is updated by the circulation of **B** around Yee lattice surface as follows:

$$\frac{\partial \mathbf{E}}{\partial t} = c \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ b_x & b_y & b_z \end{vmatrix} \begin{aligned} &= c[\ \mathbf{i}(\tfrac{\partial b_z}{\partial y} - \tfrac{\partial b_y}{\partial z}) \\ &\quad + \mathbf{j}(\tfrac{\partial b_x}{\partial z} - \tfrac{\partial b_z}{\partial x}) \\ &\quad + \mathbf{k}(\tfrac{\partial b_y}{\partial x} - \tfrac{\partial b_x}{\partial y})] \end{aligned} \tag{20}$$

Thus the electric field components e_x, e_y, e_z are, respectively, updated by the circulation of **B** around Yee lattice surface as follows:

$$\begin{aligned} \tfrac{\partial}{\partial t} e_x &= (e_x^{new}(i+.5,j,k) - e_x^{old}(i+.5,j,k))/\delta t \\ &= c[(b_z(i+.5,j+.5,k) - b_z(i+.5,j-.5,k))/\delta y \\ &- (b_y(i+.5,j,k+.5) - b_y(i+.5,j,k-.5))/\delta z], \end{aligned} \tag{21}$$

$$\begin{aligned} e_x^{new}(i,j,k) &= e_x^{old}(i,j,k) \\ &+c[\ b_y(i,j,k-1) - b_y(i,j,k) - b_z(i,j-1,k) + b_z(i,j,k)\ ], \end{aligned} \tag{22}$$

$$\begin{aligned} \tfrac{\partial}{\partial t} e_y &= (e_y^{new}(i,j+.5,k) - e_y^{old}(i,j+.5,k))/\delta t \\ &= c[(b_x(i,j+.5,k+.5) - b_x(i,j+.5,k-.5))/\delta z \\ &- (b_z(i+.5,j+.5,k) - b_z(i-.5,j+.5,k))/\delta x], \end{aligned} \tag{23}$$

$$\begin{aligned} e_y^{new}(i,j,k) &= e_y^{old}(i,j,k) \\ &+c[\ b_z(i-1,j,k) - b_z(i,j,k) - b_x(i,j,k-1) + b_x(i,j,k)\ ], \end{aligned} \tag{24}$$

$$\begin{aligned}\frac{\partial}{\partial t}e_z &= (e_z^{new}(i,j,k+.5) - e_z^{old}(i,j,k+.5))/\delta t \\ &= c[\ (b_y(i+.5,j,k+.5) - b_y(i-.5,j,k+.5))/\delta x \\ &\quad - (b_x(i,j+.5,k+.5) - b_x(i,j-.5,k+.5))/\delta y],\end{aligned} \tag{25}$$

$$\begin{aligned}e_z^{new}(i,j,k) &= e_z^{old}(i,j,k) \\ &\quad +c[\ b_x(i,j-1,k) - b_x(i,j,k) - b_y(i-1,j,k) + b_y(i,j,k)].\end{aligned} \tag{26}$$

After updating the electric field by the circulation of the magnetic field around that Yee lattice surface, charge flux **J** are calculated and subtracted after the particles are moved later in the program.

2.4 Particle Update

Newton-Lorentz equations are already in typical "update" form. The time centered finite difference version of the Newton-Lorentz particle update is:

$$\mathbf{v}^{new} - \mathbf{v}^{old} = \frac{q\delta t}{m} < \mathbf{E} + \frac{1}{2}(\mathbf{v}^{new} + \mathbf{v}^{old}) \times \mathbf{B} > \tag{27}$$

$$\mathbf{r}^{next} - \mathbf{r}^{present} = \delta t \mathbf{v}^{new} \tag{28}$$

This shows that position must be leap-frogged over velocities. Hatree and Boris found a good physical interpretation of the steps in this explicit procedure:

[1] Half an electric acceleration:

$$\mathbf{v}_0 \longleftarrow \mathbf{v}^{old} \tag{29}$$

or

$$\mathbf{v}_0 = \mathbf{v}^{old} + q\mathbf{E}\delta t/2m \tag{30}$$

[2] Pure magnetic rotation:

$$\mathbf{v}_1 \longleftarrow \mathbf{v}_0 \tag{31}$$

or

$$\mathbf{v}_1 - \mathbf{v}_0 = (\mathbf{v}_1 + \mathbf{v}_0) \times q\mathbf{B}\delta t/2m \tag{32}$$

[3] Another half electric acceleration:

$$\mathbf{v}^{new} \longleftarrow \mathbf{v}_1 \tag{33}$$

or

$$\mathbf{v}^{new} = \mathbf{v}_1 + q\mathbf{E}\delta t/2m \tag{34}$$

The Eq. (31) determining $\mathbf{v}_1$ from $\mathbf{v}_0$ is still implicit but its explicit form follows from: (1) dotting with $\mathbf{v_1} + \mathbf{v_0}$ to check that the magnetic field does not work and that the magnitudes of $\mathbf{v}_1$ and $\mathbf{v}_0$ are the same, (2) dotting with **B** to check that components along **B** are the same, (3) crossing with $q\mathbf{B}\delta/2m$ and substituting back, then to give

$$\mathbf{v}_1 = \mathbf{v}_0 + 2 \times \frac{\mathbf{v_0} + \mathbf{v_0} \times \mathbf{b_0}}{1 + b_0^2} \times \mathbf{b}_0 \tag{35}$$

2.5 Relativistic Generalization

In the code, the particle trajectory is integrated using a time-centered leap-frog scheme. Let

$$\mathbf{u} = \mathbf{v} \ , \qquad \gamma^2 = (1 - \frac{u^2}{c^2})^{-1} \tag{36}$$

Here γ is denoted by relativistic factor. Newton-Lorentz Eq. (27) gives:

$$\mathbf{u}^{n+\frac{1}{2}} - \mathbf{u}^{n-\frac{1}{2}} = \frac{q\delta t}{m}[\mathbf{E}^n + \frac{1}{2\gamma^n}(\mathbf{u}^{n+\frac{1}{2}} + \mathbf{u}^{n-\frac{1}{2}}) \times \mathbf{B}^n] \tag{37}$$

$$\mathbf{r}^{n+1} = \mathbf{r}^n + \mathbf{v}^{n+\frac{1}{2}}\delta t = \mathbf{r}^n + \frac{\mathbf{u}^{n+\frac{1}{2}}\delta t}{\gamma^{n+\frac{1}{2}}} \tag{38}$$

where

$$(\gamma^{n+\frac{1}{2}})^2 = 1 + (\frac{u^{n+\frac{1}{2}}}{c})^2 \tag{39}$$

2.6 Force Interpretations

In Eq. (37), **E** and **B** are interpolated from the grids to the particle positions in Yee lattice. Throughout the code, linear interpolation is employed for subgrid resolution. This means that there is no stringent lower limit to the sizes of such quantities as gyroradii or Debye lengths. For quantities recorded on the integer mesh $x = i$, $y = j$, $z = k$, this means interpolating the eight nearest entries by applying weights so-called "volume" weights (Buneman, 1993). For example, the "volume" weight for (i, j, k) is $(1 - dx)(1 - dy)(1 - dz) = cx * cy * cz$ and for $(i + 1, j + 1, k + 1)$ is $dx * dy * dz$.

In Yee lattice as shown in Fig. 4, the interpolated force at (x, j, k) exerted by the electric field components e_x is denoted by $\mathbf{F}_{e_x}^{(x,j,k)}$ and expressed as follows:

$$\mathbf{F}_{e_x}^{(x,j,k)} = e_x(i, j, k) + [e_x(i + 1, j, k) - e_x(i, j, k)]\delta x, \tag{40}$$

Here

$$e_x(i, j, k) = \frac{1}{2}\{e_x(i, j, k) + e_x(i - 1, j, k)\}, \tag{41}$$

$$e_x(i + 1, j, k) = \frac{1}{2}\{e_x(i + 1, j, k) + e_x(i, j, k)\}. \tag{42}$$

In Yee lattice, please note that the electric field and magnetic field are staggered as shown in the Fig. 4. Thus we obtain

$$\begin{aligned} 2\mathbf{F}_{e_x}^{(x,j,k)} &= e_x(i, j, k) + e_x(i - 1, j, k) \\ &+ [\ e_x(i + 1, j, k) - e_x(i - 1, j, k)]\delta x. \end{aligned} \tag{43}$$

The interpolated forces exerted by e_x at $(x, j+1, k)$, $(x, j, k+1)$, and $(x, j+1, k+1)$ are

$$\begin{aligned} 2\mathbf{F}_{e_x}^{(x,j+1,k)} &= e_x(i, j + 1, k) + e_x(i - 1, j + 1, k) \\ &+ [\ e_x(i + 1, j + 1, k) - e_x(i - 1, j + 1, k)]\delta x, \end{aligned} \tag{44}$$

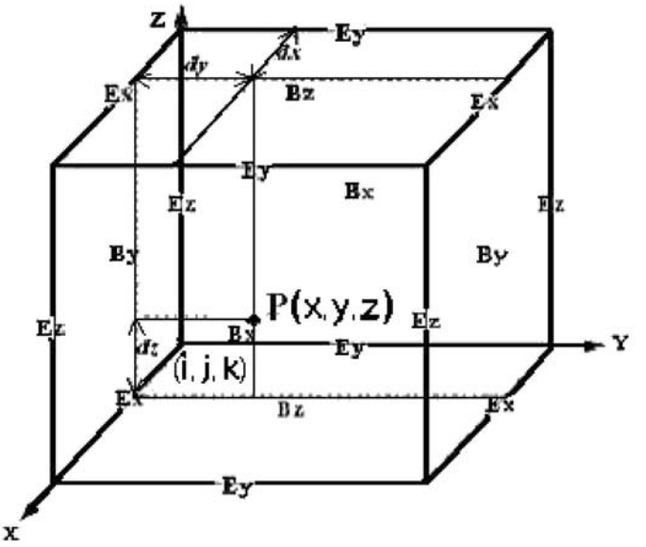


Fig. 4. The positions of field components in Yee lattice. A particle is located at the point P.

$$\begin{aligned} 2\mathbf{F}_{e_x}^{(x,j,k+1)} &= e_x(i,j,k+1) + e_x(i-1,j,k+1) \\ &+ [\, e_x(i+1,j,k+1) - e_x(i-1,j,k+1)]\delta x, \end{aligned} \tag{45}$$

and

$$\begin{aligned} 2\mathbf{F}_{e_x}^{(x,j+1,k+1)} &= e_x(i,j+1,k+1) + e_x(i-1,j+1,k+1) \\ &+ [\, e_x(i+1,j+1,k+1) - e_x(i-1,j+1,k+1)]\delta x. \end{aligned} \tag{46}$$

respectively. Thus the interpolated forces exerted by e_x at (x, y, k), $(x, y, k+1)$, and (x, y, z), are

$$\mathbf{F}_{e_x}^{(x,y,k)} = \mathbf{F}_{e_x}^{(x,j,k)} + [\mathbf{F}_{e_x}^{(x,j+1,k)} - \mathbf{F}_{e_x}^{(x,j,k)}]\delta y, \tag{47}$$

$$\mathbf{F}_{e_x}^{(x,y,k+1)} = \mathbf{F}_{e_x}^{(x,j,k+1)} + [\mathbf{F}_{e_x}^{(x,j+1,k+1)} - \mathbf{F}_{e_x}^{(x,j,k+1)}]\delta y, \tag{48}$$

and

$$\mathbf{F}_{e_x}^{(x,y,z)} = \mathbf{F}_{e_x}^{(x,y,k)} + [\mathbf{F}_{e_x}^{(x,y,k+1)} - \mathbf{F}_{e_x}^{(x,y,k)}]\delta z. \tag{49}$$

respectively. The interpolated forces $\mathbf{F}_{e_y}^{(x,y,z)}$, $\mathbf{F}_{e_z}^{(x,y,z)}$, $\mathbf{F}_{b_x}^{(x,y,z)}$, $\mathbf{F}_{b_y}^{(x,y,z)}$, and $\mathbf{F}_{b_z}^{(x,y,z)}$ exerted by the electric field components e_y, e_z, the magnetic field component b_x, b_y, and b_z can be interpolated in the same manner, respectively.

2.7 Current Deposit

As we discussed in the electric field update section, first, the change of **E** flux through a cell surface (offset grid) equals the circulation of **B** around that surface.

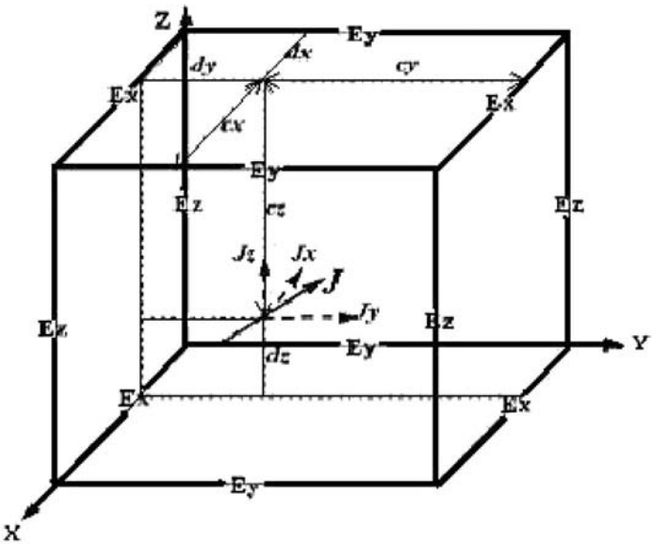


Fig. 5. The current components recorded at the point P in the Yee lattice.

Then the charge fluxes are subtracted from the **B** circulation later. TRISTAN does not employ a charge density array. Only charge fluxes, i. e., the amounts of charge flowing the faces of Yee lattice, are needed. From the Maxwell equations, one notes that Poisson equation will always be satisfied if the charge conservation condition:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{J} \tag{50}$$

is satisfied. Hence the electromagnetic field can be updated from only two curl Maxwell equations if one can enforce rigorous charge conservation numerically. A rigorous charge conservation method for current deposit is described in detail in (Villasenor and Buneman, 1992). In this scheme, one obtains the current flux through every cell surface within a time step δt by counting the amount of charge carried across the Yee lattice cell surfaces by particles as they move from $\mathbf{r^n}$ to $\mathbf{r^{n+1}}$ as shown in Fig. 5. In Yee lattice cell surfaces, the charge fluxes are subtracted from each component of **E** field as follows:

$$\begin{aligned}
e_x(i,j,k) &= e_x(i+.5,j,k) \\
&= e_x(i,j,k) - J_x * cy * cz, \\
e_x(i,j+1,k) &= e_x(i+.5,j+1,k) \\
&= e_x(i,j+1,k) - J_x * dy * cz, \\
\\
e_x(i,j,k+1) &= e_x(i+.5,j,k+1) \\
&= e_x(i,j,k+1) - J_x * cy * dz, \\
e_x(i,j+1,k+1) &= e_x(i+.5,j+1,k+1) \\
&= e_x(i,j+1,k+1) - J_x * dy * dz,
\end{aligned}$$

$$
\begin{aligned}
e_y(i,j,k) &= e_y(i,j+.5,k) \\
&= e_y(i,j,k) - J_y * cx * cz, \\
e_y(i+1,j,k) &= e_y(i+1,j+.5,k) \\
&= e_y(i+1,j,k) - J_y * dx * cz, \\
e_y(i,j,k+1) &= e_y(i,j+.5,k+1) \\
&= e_y(i,j,k+1) - J_y * cx * dz, \\
e_y(i+1,j,k+1) &= e_y(i+1,j+.5,k+1) \\
&= e_y(i+1,j,k+1) - J_y * dx * dz.
\end{aligned}
$$

and

$$
\begin{aligned}
e_z(i,j,k) &= e_z(i,j,k+.5) \\
&= e_z(i,j,k) - J_z * cy * cx, \\
e_z(i,j+1,k) &= e_z(i,j+1,k+.5) \\
&= e_z(i,j+1,k) - J_z * dy * cx, \\
e_z(i+1,j,k) &= e_z(i+1,j,k+.5) \\
&= e_z(i+1,j,k) - J_z * cy * dx, \\
e_z(i+1,j+1,k) &= e_z(i+1,j+1,k+.5) \\
&= e_z(i+1,j+1,k) - J_z * dy * dx.
\end{aligned}
$$

2.8 Position of Magnetopause

In original TRISTAN code, the Earth dipole field is located inside the simulation domain. Before running simulations, the rough size of the Earth magnetosphere should be determined and the size should be small enough to be inside the simulation domain.

As shown in Fig. 6, the Ampere equation gives

$$\mathbf{B} = \frac{\mu_0}{4\pi} \int \frac{I_0 d\mathbf{l} \times \mathbf{r}}{r^3} \tag{51}$$

where the small vector element of the ring is $d\mathbf{l} = -\mathbf{e}_\phi r_0 d\phi$, the distance from arbitrary point along the y-axis to $d\mathbf{l}$ is $r^2 = R^2 + r_0^2 - 2Rr_0^2 \cos\phi$, and the ring current is I_0. Here r_0 is the ring radius, ϕ is the ring angle. From $R^2 = r^2 + r_0^2 - 2rr_0 \cos(\alpha + \pi/2)$, we get:

$$\sin\alpha = \frac{R\cos\phi - r_0}{r}. \tag{52}$$

Thus the B_z at $(0, R, 0)$ is:

$$B_z = -\frac{\mu_0 I_0 r_0}{4\pi} \int \frac{R\cos\phi - r_0}{r^3} d\phi \tag{53}$$

when $R \gg r_0$ is assumed, it can be

$$B_z(\text{magnetopause}) \sim \frac{\mu_0 I_0 r_0^2}{2R^3} \tag{54}$$

The potential energy density of the Earth magnetic field in magnetopause is:

$$W_{magnetic} = B^2/2\mu_0 \sim B_z^2/2\mu_0 \sim \frac{\mu_0 I_0^2 r_0^4}{8R^6}. \tag{55}$$

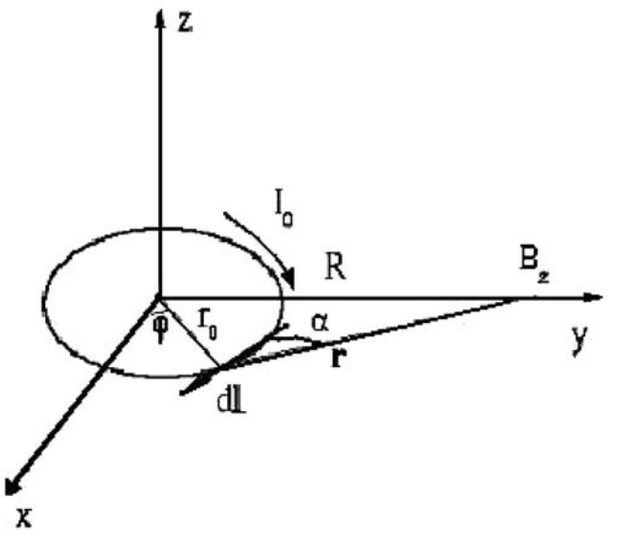


Fig. 6. The ring current to generate the dipole field of the Earth at $(0, R, 0)$.

The kinetic energy density of solar wind flow is:

$$W_{\text{solar wind}} = \frac{1}{2}\rho_{solar\ wind}V^2_{solar\ wind} \sim \frac{1}{2}(m_i D_i + m_e D_e)V^2_{\text{solar wind}}. \tag{56}$$

where D_i, D_e are the number density of ions and electron in solar wind respectively. The position of the magnetopause, R_{MP}, located where the two energy densities are equal to each other. We obtain it from Eq. (55) and Eq. (56):

$$R^6_{MP} \sim \frac{\mu_0 I_0^2 r_0^4}{4(m_i D_i + m_e D_e)V^2_{\text{solar wind}}} \tag{57}$$

In TRISTAN code, the Earth dipole filed is ramped-up linearly at an initial stage of the simulation. The half dipole field ramp-up time is $\frac{-o2}{o3} = n$, and the final ring current charge density is $o_{final} = -o3 \times n(n+1)(2n+1)/3$. The ratio -o2/o3 should be an exact integer n. The number of time steps taken for the build-up is $2n$. The final value of "o" will be $-o3n(n+1)(2n+1)/3$ in the simulation. Using Eq. (57) and the final value of "o", the rough location of the magnetopause can be estimated.

2.9 Field Densities

It is most important to realize that the three components of each field vector have not been recorded at the same places in Yee lattice. To get the electric and magnetic field components at location (i, j, k), one must form the following averages;

$$e_x(i, j, k) = \frac{1}{2}\{e_x(i-1, j, k) + e_x(i, j, k)\}, \tag{58}$$

$$e_y(i,j,k) = \frac{1}{2}\{e_y(i,j-1,k) + e_y(i,j,k)\}, \tag{59}$$

$$e_z(i,j,k) = \frac{1}{2}\{e_z(i,j,k-1) + e_z(i,j,k)\}, \tag{60}$$

$$\begin{aligned} b_x(i,j,k) = \ & \tfrac{1}{2}\{\ \tfrac{1}{2}\ [\ b_x(i,j,k) + b_x(i,j,k-1)\ ] \\ & +\tfrac{1}{2}\ [\ b_x(i,j-1,k) + b_x(i,j-1,k-1)\ ]\}, \end{aligned} \tag{61}$$

$$\begin{aligned} b_y(i,j,k) = \ & \tfrac{1}{2}\{\ \tfrac{1}{2}\ [\ b_y(i,j,k) + b_y(i-1,j,k)\ ] \\ & +\tfrac{1}{2}\ [\ b_y(i,j,k-1) + b_y(i-1,j,k-1)\ ]\}, \end{aligned} \tag{62}$$

and

$$\begin{aligned} b_z(i,j,k) = \ & \tfrac{1}{2}\{\ \tfrac{1}{2}\ [\ b_z(i,j,k) + b_z(i,j-1,k)\ ] \\ & +\tfrac{1}{2}\ [\ b_z(i-1,j,k) + b_z(i-1,j-1,k)\ ]\}. \end{aligned} \tag{63}$$

2.10 Formulae and Normalization

TRISTAN code uses scales such that $\epsilon_0 = 1$ and hence $\mu_0 = 1/c^2$ (means $\mathbf{E} = \mathbf{D}$). TRISTAN also uses scales such that time step $\delta t = 1$, grid sizes $\delta x = \delta y = \delta z = 1$, electron charge to mass ratio $q_e/m_e = -1$, electron mass $m_e = 1$. Thus the normalized ion and electron cyclotron frequencies are:

$$\Omega_e = \frac{zeB}{m_e} = \frac{B}{m_e} = B, \tag{64}$$

$$\Omega_i = \frac{zeB}{m_i} = \frac{B}{m_i} = B \times \mathrm{rmass}. \tag{65}$$

respectively. Here the mass ration $\mathrm{rmass} = m_e/m_i$. The normalized ion and electron gyroradii are:

$$\rho_e = \frac{m_e v_e}{zeB} = \frac{m_e v_e}{B} = \frac{v_e}{B}, \tag{66}$$

$$\rho_i = \frac{m_i v_i}{zeB} = \frac{m_i v_i}{B} = \frac{v_i}{\mathrm{rmass}B}. \tag{67}$$

respectively. The normalized electron and ion plasma frequencies are:

$$\omega_{pe} = \sqrt{\frac{n_e z^2 e^2}{\epsilon_0 m_e}} = \sqrt{\frac{n_e}{m_e}} = \sqrt{n_e}, \tag{68}$$

$$\omega_{pi} = \sqrt{\frac{n_i z^2 e^2}{\epsilon_0 m_i}} = \sqrt{\frac{n_i}{m_i}} = \sqrt{n_i \mathrm{rmass}}. \tag{69}$$

respectively. The normalized Alfven wave speed is:

$$V_A = \sqrt{\frac{B_0^2}{\mu_0 \rho_m}} = c\sqrt{\frac{B_0^2}{\rho_m}}. \tag{70}$$

where $\rho_m = n_e m_e + n_i m_i = m_e(n_e + \frac{n_i}{\text{rmass}})$ is the mass density. The normalized sound speed is:

$$C_s = \sqrt{\frac{\gamma_e T_e + \gamma_i T_i}{m_i}} = \sqrt{\frac{\gamma p_0}{\rho_0}} \sim \sqrt{\frac{\gamma T}{m_i}}$$

$$\sim \sqrt{\frac{T_e}{m_i}} = \sqrt{\frac{\text{rmass } T_e}{m_e}} = \sqrt{\text{rmass } T_e}. \tag{71}$$

The normalized Debye length is:

$$\lambda_{De} = \sqrt{\frac{\epsilon_0 T_e}{n_e e^2}} = \sqrt{\frac{T_e}{n_e}}. \tag{72}$$

In TRISTAN code, users have to specify the solar wind velocity $vdrft$, the thermal ion velocity $vth2i$, the thermal electron velocity $vth2e$, the particle density D_{pair}, the half dipole field ramp-up time $\frac{o2}{o3} = N$, and the final ring current charge density:

$$o_{final} = -o3 \times N(N+1)(2N+1)/3. \tag{73}$$

From Eqs. (68), (69) and (72), please note that the plasma frequency is proportional to ${n_e}^{\frac{1}{2}}$ and the Debye length is proportional to ${n_e}^{-\frac{1}{2}}$. Thus, fixing all input parameters except the particle density, if we change the particle density in the simulation to reduce the statistical noises or fluctuations, all physical quantities will vary and the physical meanings of simulation results will change at the same time. Varying the value of the particle density, we have to adjust all other physical input parameters simultaneously to keep the physical problems fixed.

2.11 Stabilities and Heating Conditions

Solving Maxwell equations by the centered difference scheme in space and by leap-frog method in time, the spatial grid $\delta x, \delta y, \delta z$ and the time step δt should satisfy the following inequality, which is called Courant condition,

$$\delta x, \delta y, \delta z > c\, \delta t \tag{74}$$

where c is the light speed and $\delta x, \delta y, \delta z$ is the grid size. The condition is easily derived from the numerical dispersion relation of the light mode.

In TRISTAN code, if we consider the real physical system size of the Earth magnetosphere, it is definitely impossible to use the grid size equal to or more than one Debye length even if we use a very powerful parallel computer. You can imagine that the typical Debye length in magnetopause is roughly an order of 10 m and how many grids will be needed to simulate the whole magnetosphere. One trade-off to solve this problem is to reduce the grid size less than one Debye length. However, we have to carefully avoid nonphysical instabilities caused by the grid or numerical grid heating (Birdsall and Langdon, 1985). For a Maxwellian velocity distribution with no drift, a rough rule of thumb is that nonphysical instability has ignorable growth for $\lambda_{De}/\delta x > \frac{1}{\pi} \sim 0.3$ for linear

weighting. However, when $\lambda_{De} \sim 0.1\delta x$, the lowest and strongest aliases interact with the steep sides of a Maxwellian velocity distribution and there is little Landau damping. The result is strong numerical instability. If $\lambda_{De}/\delta x$ decreases further the instability goes away, as it should since a cold stationary plasma is inactive. Therefore, when the Debye length λ_{De} is determined, a very rough criteria to avoid nonphysical instability is to avoid the range:

$$o(10^{-1}) \leq \frac{\lambda_{De}}{\delta x} \leq 0.3.$$

Of course, there are many other types of nonphysical instabilities and we have to check them carefully in the simulation.

3 Arrays in Original TRISTAN Code

The motivation of TRISTAN, a fully three-dimensional (3D) electromagnetic (EM) particle-in-cell (PIC) code written by Oscar Buneman and other collaborators in Stanford University, is to develop a general particle-in-cell code for space plasma simulations (Buneman, 1993). In this section, we only discuss the data structure and the data distribution over processors on the HPF TRISTAN code. For the detail physics of the PIC code in general, please refer to, for examples, (Birdsall and Langdon, 1985) and (Walker, 1991).

The data structure of TRISTAN code consists of two primitive data types. The first one is the particle data as follows:

$$x(mp), y(mp), z(mp), u(mp), v(mp), w(mp),$$

where mp= total number of particles, the positions and velocities of ions and electrons are recorded at:

$$x(1:mh), y(1:mh), z(1:mh),$$

$$u(1:mh), v(1:mh), w(1:mh),$$

and

$$x(mh+1:mp), y(mh+1:mp), z(mh+1:mp),$$

$$u(mh+1:mp), v(mh+1:mp), w(mh+1:mp).$$

respectively, where $mh = mp/2$. The second one is the grided field data expressed as the triple-indexed arrays of EM (ElectroMagnetic) fields as follows:

$$ex(i,j,k), ey(i,j,k), ez(i,j,k),$$

and

$$bx(i,j,k), by(i,j,k), bz(i,j,k).$$

The original TRISTAN code uses "COMMON" block clause to save and transfer fields data between subroutines in the **MOVER** (push particles) and **DEPOSIT** (deposit current data to the field grids) subroutine calls. Meanwhile

in the subroutines that processes the surfaces and edges of the grid data, the filed data are transferred by dummy arrays in the original code. In both of these subroutines, the field arrays are treated as single-indexed. On the other hand, triple-indexed field arrays are employed in the field solver subroutines. In the code, single-indexed arrays are converted automatically to the triple-indexed arrays when they passed over two subroutines.

Converting a serial Fortran program to a HPF program, we have to stress two points that are very important for rewriting TRISTAN in HPF: [**1**] The "COMMON" statement is restricted as suggested by pgHPF user guide and there they indicated 'We strongly recommended that programmers writing new F90 code use features like "MODULE"... to avoid the use of "COMMON"...' (Koelbel, et al., 1994, Foster, 1995), in case of data overlapping, and substituted it by "MODULE" block; [**2**] To control the communications, all the arrays are treated as fixed indexes throughout the whole program. We control the communication parts using both the "field manager" and "particle manager" (Decyk, 1995).

4 Field Data Domain Decompositions

The field data are decomposed over sub-domains of that number is equal to the number of the processors used in the simulation as indicated in Fig. 1. In processing the current deposition that is so-called the scatter part of the computations, to avoid large transients or variations of currents, TRISTAN uses a "smoother" that has 27 different weights, smoothing the current deposition. In **DEPOSIT** subroutine the smoothing is performed as follows:

$$\begin{aligned} ey(i+smx+1, j+smy, k+smz+1, Np) = \\ ey(i+smx+1, j+smy, k+smz+1, Np) \\ -sv*dz*dx-ss \end{aligned} \tag{75}$$

where $smx = -1:1, smy = -1:1, smz = -1:1$,$sv = sm(smx, smy, smz, Np)*$ qv and $ss = sm(amx, amy, amz, Np) * delt$. For details, see p. 73 and p. 321 of Lecture note by Buneman (1993). (Note that one dimensional array for ex, ey, ez is used.) Therefore, the current deposition of one particle will be related to three grids in each dimension, where one of them are at the backward grid and another at the forward grids in each dimension.

In the "MODULE" block, the field arrays are written in HPF directives as follows:

$$\mathbf{REAL}, \mathbf{DIMENSION}(nx, j, k, Np) :: ex, ey, ez$$

$$\mathbf{REAL}, \mathbf{DIMENSION}(nx, j, k, Np) :: bx, by, bz$$

where Np is the number of processor, $nx = i/Np + 3$ (here assuming i/Np is not necessarily equal to be integer exactly) keeping one guard cell in the left (backward) and right (forward) sides of the sub-domains in the domain-decomposition direction (i.e., in the solar-magnetotail direction). Here the indices i, j and k correspond to the numbers of field grids in x, y and z directions,

respectively. Using the HPF directive "DISTRIBUTE", we, respectively, map the sub-domains to each processor on a distributed parallel computer:

DISTRIBUTE(*,*,*,BLOCK) ONTO Np:: ex,ey,ez

DISTRIBUTE(*,*,*,BLOCK) ONTO Np:: bx,by,bz

In order to separate the communication parts from the computation parts, each sub-domain keeps extra cells, the so-called guard or ghost cells, that store the field data information in the first and last grids of that sub-domain in the decomposition direction. Fig. 1 illustrates this concept of the data mapping over the sub-domains or processors. Here the communications are required after updating the field data every time step. In the field manager (Decyk, 1995), the data sent to the neighbor processors are packed in the working arrays: $Cex(1, j, k, Np)$, $Cey(1, j, k, Np)$, and $Cez(1, j, k, Np)$, before they are sent to the neighbor sub-domains. Thus the field data communications are performed by the pgHPF $CSHIFT$ construct after the data are packed in the working arrays. The followings are the related parts of the HPF programs in the field manager (Decyk, 1995):

```
Cex(1,:,:,:)=ex(2,:,:,:)

Cex=CSHIFT(Cex,+1,4)

ex(nx-1,:,:,:)=Cex(1,:,:,:)

...
```

5 Particle Data Domain Decompositions

The particle data can be written in HPF directives as follows:

$$\mathbf{REAL}, \mathbf{DIMENSION}(m, Np) :: xe, ye, ze, xi, yi, zi$$

$$\mathbf{REAL}, \mathbf{DIMENSION}(m, Np) :: ue, ve, we, ui, vi, wi$$

where i and e, respectively, stand for ion and electron, the number m is the array size in each sub-domain. To ensure that the enough space are reserved to store the particle data due to the load-imbalance, m must be 10-30 % larger than the average number of particles. The number Np is the number of processors, and is the index used in the HPF "DISTRIBUTE" directive. As the particles move in time in the simulations, the physical position of some particles may cross the sub-domain boundaries, and move to the neighbor sub-domains. When a particle moves from one sub-domain to another, the data of the particle left the sub-domain must be sent to the appropriate neighbor processor at every time step. Before updating and sending the particle data, we have to sort the particles that should be sent to another sub-domain, and pack them in the working

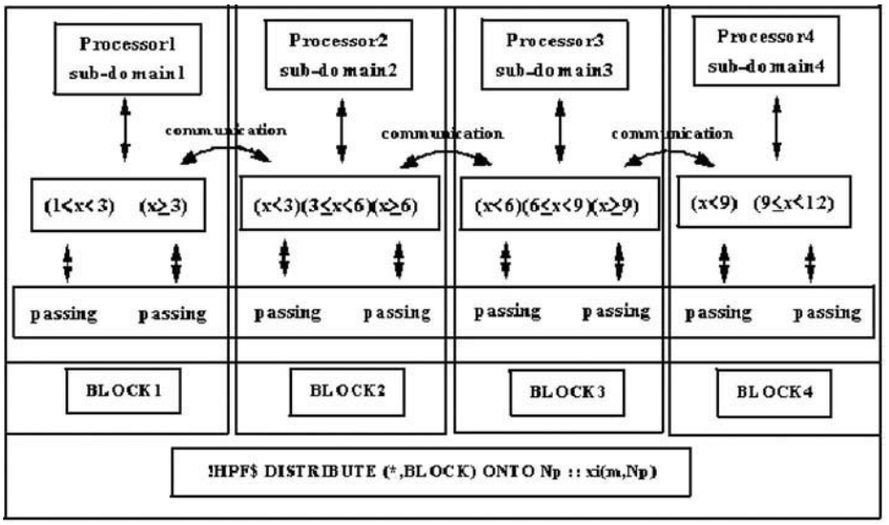


Fig. 7. Diagram of particle array decompositions and communications, with processor number $Np = 4$, grid number in decomposition direction=12.

arrays: $CRi(:, Np)$, $CLi(:, Np)$, $CRe(:, Np)$, and $CLe(:, Np)$. The number of the ions and electrons sent in right and left are denoted by the arrays $ionspsR(Np)$, $ionspsL(Np)$, $lecspsR(Np)$, and $lecspsL(Np)$, respectively. In our HPF TRISTAN code, we send both the packed arrays and their particles number arrays to the neighbor sub-domains as follows:

```
CRi=CSHIFT(CRi, -1,2)

ionspsR=CSHIFT(ionspsR, -1)

...
```

Fig. 7 shows the example of the particle data distributions and communications. After both the particle numbers and the packed working arrays are sent and received by each appropriate processors, the received particles are sorted and put into the appropriate part of the particle arrays in that sub-domain. The communications and sorting of these particles are performed in the particle manager (Decyk, 1995) as shown in Fig. 2.

6 Programming Comments on HPF Communications in PC Cluster

For benchmark purpose of HPF TRISTAN, a dual PentiumPro PC cluster consists of 16 PCs and each PC have dual 200MHz PentinumPros with 128MB EDD

DIMM memories. The PCs in the PC cluster system are hooked through 100 Base-T ethernet with 100 Base-T switching Hub. Redhat Linux version 4.1 was used as their operating systems. The pgHPF compiler version 1.7 was installed for HPF computations.

One of the most difficult problems in our HPF TRISTAN code is the communication programming, especially, the determination of the buffer sizes which is used to pack the data sent to the neighbor processors. Of course, we can define a buffer size large enough to send the particle or grid data to neighbor processors at one time. However, as shown in Fig. 8, our experience shows that when the buffer sizes become larger than some critical values, in this case 1456 bytes in our PC cluster system, the communication suddenly becomes unstable, and the communication times suddenly jump up to 5 to 8 times larger than those less than the critical value 1456 bytes. As indicated in the Fig. 8, the communication times when the buffer size go beyond 1456 bytes are not uniquely determined and rather indeterministic. In order to avoid the sudden communication slow-down, we have to carefully choose the buffer size. We have to split the particles or grids data into smaller pieces of buffers, pack the smaller data, and send the data to the neighbor processors one by one. Thus we can avoid the large slow-down of the simulations in this system due to the unstable HPF communications. In our HPF TRISTAN code, the buffer sizes can be varied and can be set without modifying the program. We can first evaluate the best buffer size and run the simulations. The best buffer size can be chosen as indicated in Fig. 8. For ex-

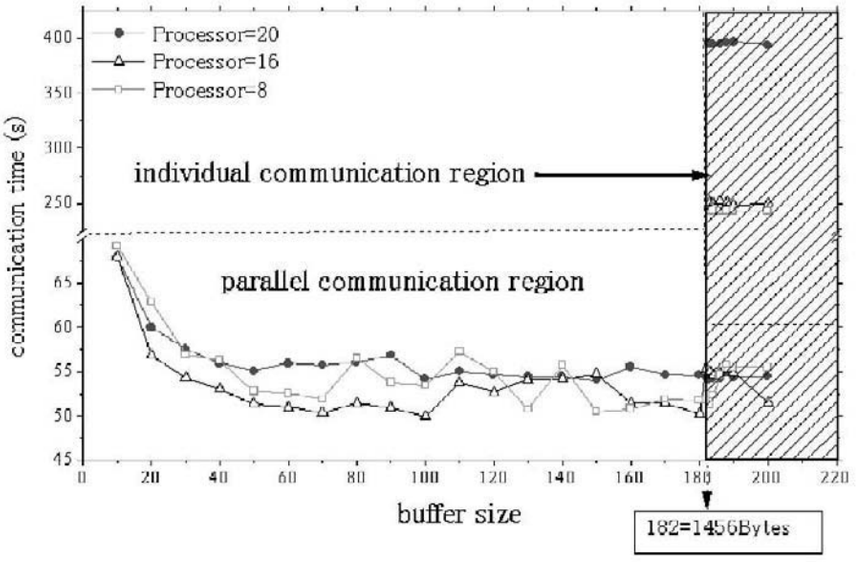


Fig. 8. Buffer sizes and CSHIFT communication times in HPF.

amples, in this figure, the best buffer sizes can be chosen between 640 to 1400 bytes.

The reasons for performance degradation in communications with this longer packets than 1456 bytes are not yet investigated in detail. One possibility of this degradation is due to MTU of the ethernet. MTU is the Maximum Transmission Unit that IP is allowed to use for a particular interface. If your MTU is set too big, in this case beyond ∼1456 bytes your packets must be fragmented, or broken up, by a switching hub along the path to the other PCs. This may result in a drastic decrease in throughput. However, we have not identified the source of this communication degradation. We would like to leave this investigation to our future research.

7 Benchmark and Simulation Results

In Table 1, the parameter $\varepsilon_{eff-grid}$ is defined as:

$$\varepsilon_{eff-grid} = \frac{Num_{total\ grid} - Num_{total\ guard\ cell}}{Num_{total\ grid}}.$$

where $Num_{total\ grid}$ is the total grid number in decomposition direction, and $Num_{total\ guard\ cell}$ is the total number of guard cell. Table 1 shows the total times, speedups and parallel efficiency vs the number of processors. The total

Table 1. Benchmark resluts with time step=100, particle number =1200,000, and grid number $=185 \times 65 \times 65$.

Procs	Time(s)	speedup S_p	efficiency $\varepsilon(\%)$	$\varepsilon_{eff-grid}(\%)$
1	4836	1.0	100.	100
2	3706	1.3	65.2	96.9
3	2416	2.0	66.7	95.4
4	1769	2.7	68.4	93.9
5	1457	3.3	66.3	92.5
6	1195	4.0	67.4	91.1
7	1034	4.7	66.8	89.8
8	937.4	5.2	64.5	88.5
9	881.9	5.5	60.9	87.2
10	762.7	6.3	63.4	86.0
11	713.0	6.8	61.7	84.9
12	652.0	7.4	61.8	83.7
13	616.8	7.8	60.3	82.6
14	575.9	8.4	60.0	81.5
15	572.5	8.4	58.3	80.4
16	516.4	9.4	58.5	79.4
17	497.2	9.7	57.2	78.4
18	470.5	10.3	57.1	77.4
19	453.5	10.7	56.1	76.4
20	426.6	11.3	56.7	75.5

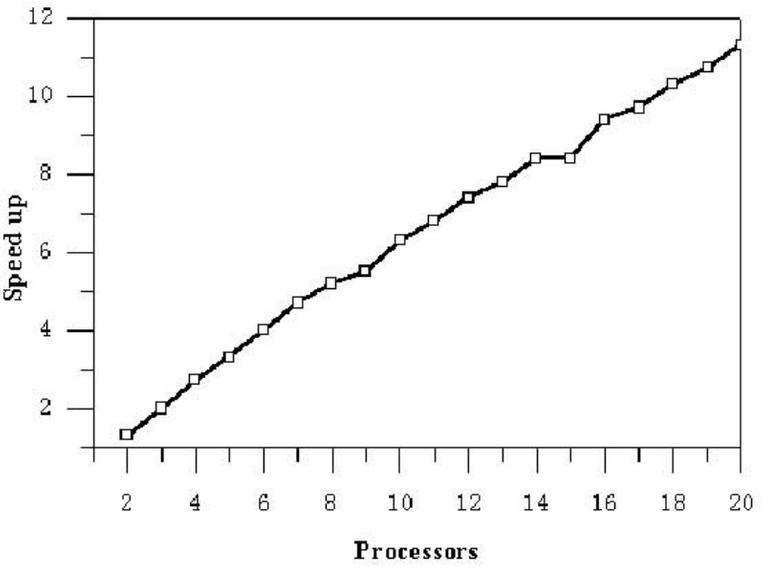


Fig. 9. Speed up vs processor number.

computation time of single processor was measured by the original version of TRISTAN code compiled by *pgf*77 compiler with the optimization level **-O2** option. Figure 9 shows the speed up vs processor number.

With fixing the problem size and increasing the processor number, the grid number in one sub-domain is reduced gradually. For example, 40 extra ghost grid cells in total must be added to each sub-domains in decomposition direction or in x for 20 processors. It is about 25 percents of the total grid number in decomposition direction in this case. Thus the communication overhead become insignificant comparing with the total PIC computation time as we increase the number of the processors. The increase of the communication overhead reduces the parallel efficiency in the table. If the communication overhead is insignificant, it is very hard to improve the parallel efficiency of the code without varying the problem size. However, even the most advanced parallel computer nowadays, it is not so easy to increase the problem sizes as we increase the number of the processors due to the large data sizes we have to store in each simulation run. Thus the optimal parallel efficiency of the scalable relation between the problem sizes and the number of processors are difficult to be measured in our simulation. However, Fig. 9 shows the high linearity of our HPF TRISTAN code and the code scales well. In addition, with the HPF compiler overhead and the load-imbalance overhead due to the Earth dipole field, the parallel efficiency around 60-65 % is affordable in this type of large scale simulations.

PIC simulations exploring the solar wind-magnetosphere interaction with this HPF code were accomplished on the PentiumPro PC cluster (Cai and Lu, 1999). After measuring the communication efficiencies via different processor

Table 2. Simulation parameters of case 1.

grids	205x165x165
initial ion-electron pairs	500,000,000
Light speed	0.5
ϵ_0	1.0
m_i / m_e	16
q_e / m_e	1.0
Δx	1
Δt	1
electron temperature T_e	0.004
ions temperature T_i	0.00025
solar-wind speed	0.25
IMF	no
ω_{pe}	0.2
$\lambda_D/\Delta x$	0.4
plasma parameters	g$\sim$2.8

numbers, the CFSHIFT function and its communication have been optimized by splitting a large data into many small size ones. So that a high performance communication is achieved. We also have run the code on Hitachi supercomputer SR2201 and Fujitsu VPP 5000 etc..

In case 1, we use a 205 by 165 by 165 grid and 50,000,000 plasma particles. The other parameters are listed in table 2: (1) the center of the loop current is located at (100Δ, 82.5Δ, 82.5Δ); (2) the solar-wind drift velocity is 0.5c, where c is the light speed; (3) $m_i/m_e = 16$; (4) use 8 particles per grid cell in average; (5) $T_e = 0.004$ and $T_i = 0.00025$; (6) the plasma parameter is about 2.8. Figure 10 (a) and (b) show the ion density profiles on the XZ and XY planes at time step 1500 respectively. It is clear that the complete configuration of magnetosphere, including bow shocks, magnetopause, magnetosheath, magnetotail, plasma sheet, and polar cusp are generated. In this case, more basic kinetic behaviors of space plasma in the magnetosphere have been investigated including a time-varying IMF. We use a northward IMF ($B_z = 0.01$) for an initial condition. At time step 500, the IMF is switched into southward ($B_z = -0.02$) at the sunward boundary of the domain. It is shown that after the arrival of a southwad IMF, due to the reconnection at dayside magnetopause the convection pattern across the entire polar cap begins to change in a few minutes. In contrast, the response of the equatorward motion of the open-closed field-line boundary that depends on the local time is delayed about 20 minutes relative to the onset of the reconnection at the dayside magnetopause. This time delay is considered as the

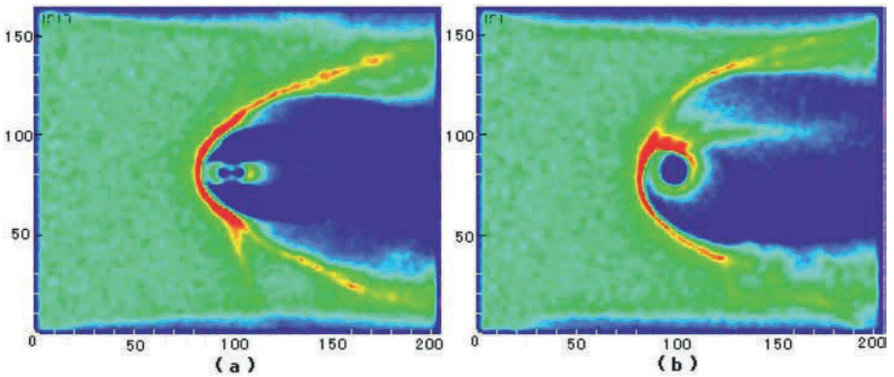


Fig. 10. The ion density profiles at (a) XZ and (b) XY planes of case 1.

Table 3. Simulation parameters of case 2.

grids	185x125x125
initial ion-electron pairs	24,000,000
Light speed	0.5
ϵ_0	1.0
m_i/m_e	16
q_e/m_e	1.0
Δx	1
Δt	1
electron temperature T_e	0.4
ions temperature T_i	0.1
solar wind speed	0.25
IMF	yes
ω_{pe}	0.2
$\lambda_D/\Delta x$	0.4
plasma parameter	g∼3

time required to convect the newly merged flux from the dayside magnetopause to the nightside inner magnetosphere.

In case 2, we use a 185 by 125 by 125 grid and 24,000,000 total particles. The other parameters are list in Table 3. It is shown in Fig. 11 that when the southward IMF arrived, the magnetic field of magnetosphere is modified. Some structures are formed. The nature of these structures is still under investigation.

We investigate the relationship between the IMF and the particle flux in polar region in case 3, in which we used a 85 by 105 by 105 grid and 3,500,000 paired particles. The other parameters are listed in Table 4. It is assumed that the cusp region is located at $25 < x < 35$, $45 < y < 60$, $56 < z < 60$. The IMF is initially zero, and switch on three times, i.e., $B_z = -0.01$ during time step 100-120, $B_z = -0.02$ during time step 600-620 and $B_z = -0.005$ during time step 900-920. The ion density profile in cusp region via time step was shown

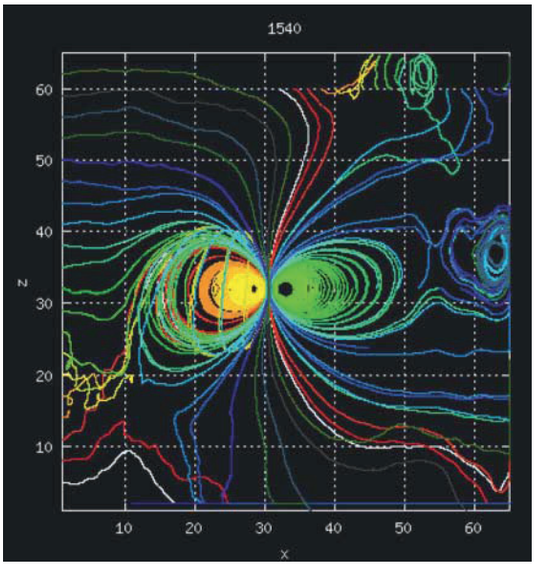


Fig. 11. Magnetic field lines at time step 1540 of case 2.

Table 4. Simulation parameters of case 3.

grids	85x105x105
initial ion-electron pairs	3,500,000
light speed	0.5
ϵ_0	1.0
m_i/m_e	16
q_e/m_e	0.5
Δx	1
Δt	1
ion temperature T_i	5.9×10^{-6}
electron temperature T_e	4.8×10^{-5}
solar wind speed	0.25
IMF	yes
ω_{pe}	0.088
$\lambda_D/\Delta x$	0.93
plasma parameter	g∼3.2

in Fig. 12. The mean ion velocities in cusp region via time step was shown in Fig. 13. The total thermal energy in cusp region via time step was shown in Fig. 14. Total particle energy in cusp region via time step was shown in Fig. 15.

It is clearly shown that the first and second switch-on of the southward IMF can cause particle flowing into cusp regions after some time steps. Many particles are entered into cusp regions. It seems that not all southward IMFs can cause

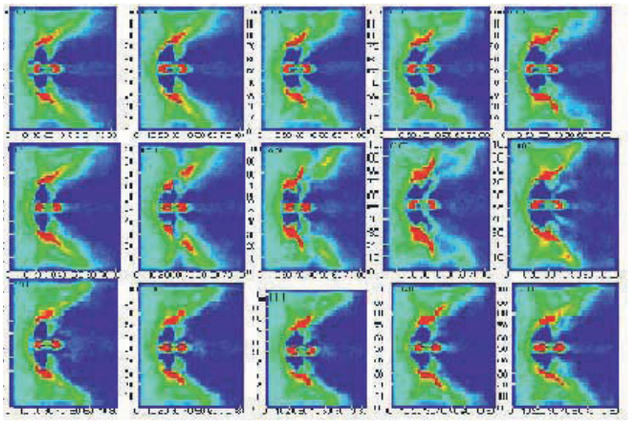


Fig. 12. The ions density profile via time step in the cusp region of case 3.

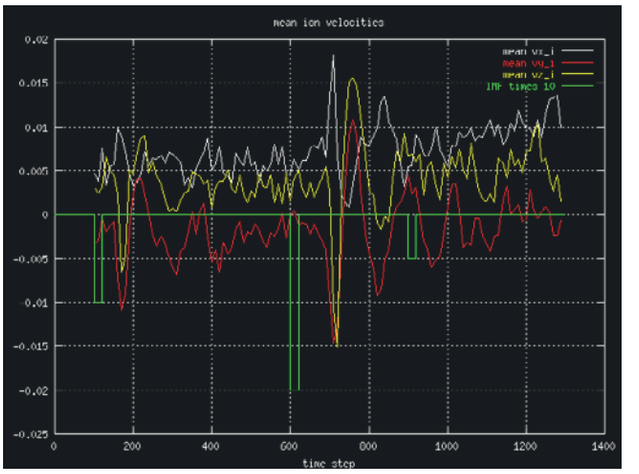


Fig. 13. Mean ion velocity and IMF via time step in cusp region of case 3.

the particle-entry into cusp regions. Only those IMFs that are strong enough can cause particle-entry into cusp regions.

The figures show that in the first two IMFs switch-on, the total particle number and the mean ion velocities decreased from the local maximum to the local minimum after the southward IMF switch-on. The intense southward IMF could push the particles into cusp region to magnetotail-ward directly. This may be related to substorm or magnetic reconnection.

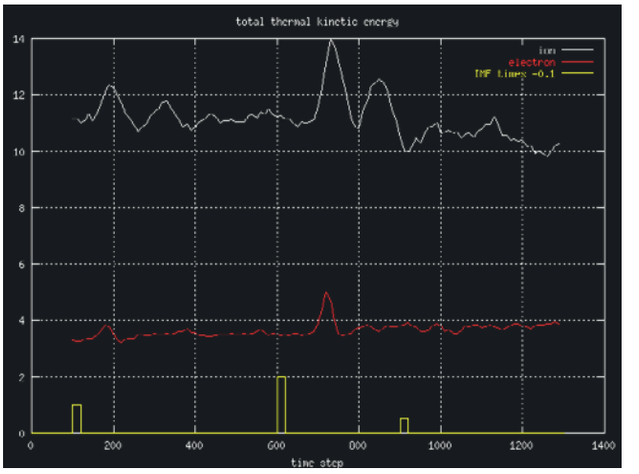


Fig. 14. Total thermal kinetic energy and IMF via time step in cusp region of case 3.

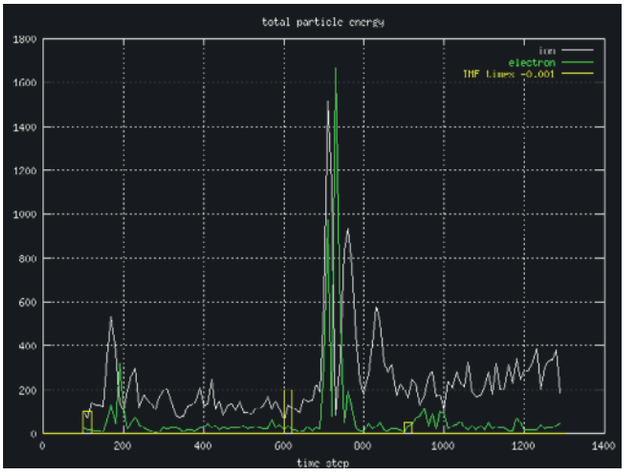


Fig. 15. Total particle energy and IMF via time step in cusp region of case 3.

8 Concluding Remarks

In the present paper, we have successfully parallelized the three-dimensional full electromagnetic and full particle code using HPF. The code is originally the same as the TRISTAN code and the code is for the space plasma simulations. As shown in Fig. 9 and Table 1, fixing the problem size, our HPF TRISTAN code has a high linearity and scales well. However, our HPF code introduces about 70% overhead and the reason for this overhead is not yet investigated. We have also parallelized the three-dimensional skeleton-PIC code introduced by V. K. Decyk (Decyk, 1995) in the same parallel algorithm (Liewer and Decyk,

1985, Decyk, 1995) using HPF. The HPF three-dimensional skeleton-PIC code introduces about 20% overheads (Cai, et al., 1999). One possibility to explain the larger overheads in our HPF TRISTAN code over the HPF skeleton-PIC code is that there are more complicated data structures in HPF TRISTAN code than those in the skeleton-PIC code. Our PCs in the cluster have no enough memory and this may degrade the performance of the PCs. Another possibility is the load-imbalance originated in the TRISTAN code as we discussed previously. Our HPF TRISTAN code has the Earth dipole filed which generates and simulates the Earth magnetosphere in one of sub-domains, and this may cause a large load-imbalance. We would like to leave the detailed investigation to our future work.

The parallelization algorithm we used in our HPF TRISTAN code is basically the same as (Liewer and Decyk, 1985) and (Decyk, 1995). We separate the communication parts from the computation parts. Thus the code can easily be converted to MPI or PVM code by replacing the HPF "CSHIFT" constructs to appropriate message passing interfaces. Our experiences show that the utilization of HPF "FORALL" or "DO INDEPENDENT" constructs in the data-parallel manner without separating the communication parts from the computation parts results in almost no gain of speedups or very poor speedups.

We have also compared the HPF skeleton-PIC code with the MPI or PVM skeleton-PIC code. The HPF code degradation of the total CPU time over the MPI or PVM code is only 10-15 % (Cai, et al., 1999) in this case. Thus we expect that we should be able to enjoy the easier HPF programming with a very small performance degradation even in the more complicate codes like the TRISTAN code.

Acknowledgment

The authors thank Professor Viktor K. Decyk for his help using the skeleton PIC codes. The authors also thank Dr. Bertrand Lembege for invaluable comments regarding to this work.

References

1. Birdsall C. K. and A.B.Langdon, Plasma Physics via Computer Simulation, McGraw-Hill, New York, (1985)
2. Buneman O., TRISTAN. In: Matsumoto H., and Omura, Y. (eds.): Computer Space Plasma Physics: Simulation Techniques and Software. Terra Scientific, Tokyo, (1993) 67-84
3. Cai, D., Q. M. Lu, and Y. T. Li, Scalability in Particle-in-Cell code using both PVM and OpenMP on PC Cluster, *Proceedings of 3rd Workshop on Advanced Parallel Processing Technologies* (1999) 69-73
4. Decyk, V. K.: Skeleton PIC Codes for Parallel Computers, Comput. Physcs. Comm. **87**, (1995) 87-94
5. Foster, I.: Designing and Building Parallel Programs, Addison-Wesley, (1995)
6. Koelbel, C.H. et al., The High Performance Fortran Handbook, The MIT press,(1994)

7. Liewer, P. C. and V. K. Decyk: A General Concurrent Algorithm for Plasma Particle-in-Cell Simulation Codes, J. Comput. Phys. **85**, (1985) 302-322
8. Homepage of Portland Group Inc.: http://www.pgroup.com
9. Villasenor, J., and O. Buneman: Rigorous Charge Conservation for Local Electromagnetic Field Solvers, Comput. Physcs. Comm. **69**, (1992) 306-316
10. Walker D. W., Particle-in-cell Plasma Simulation Codes on the Connection Machine, Computing Systems in Engineering, **1** (1991) 307-319
11. Yee, K. S, Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media, IEEE Trans. Antennas Propagat. **14**, (1966) 302-307

Full Particle Electromagnetic Simulation of Collisionless Shocks

Bertrand Lembège

CETP-CNRS-UVSQ, 10-12 Av. de l'Europe, 78140 VELIZY, France

Abstract. This tutorial-style review is mainly dedicated to the different strategies and constraints used for simulating and analysing the dynamics of collisionless shocks with full particle electromagnetic codes. The relationship between macroscopic and microscopic processes inherent to the shock dynamics and the associated numerical constraints are explained. Different initial methods are commonly used for exciting a shock which are summarized and compared in terms of computing efficiency. Boundary conditions and associated numerical "tricks" are also detailed. Advantages and limitations of using full particle codes for simulating collisionless shocks are discussed. These limitations are mainly expressed in terms of compromise by using irrealistic plasma parameters values or/and appropriate magnetic field configurations. As a consequence, such a compromise requires to validate the simulation results obtained, before being compared with real experimental results; complementary validation approaches are proposed. In the post-processing stage, different strategies (complementary to the full particle code but still based on particle analysis) are necessary when the structures and/or phenomena inherent to the shock dynamics become too intricated: they concern (i) various approaches allowing to analyse the underlying basic mechanisms, and (ii) the strategies "mimicing" real space experiments.

1 Introduction

Different strategies can be followed for analysing a collisionless shock with full particle mono and/or multi-dimensional codes. These are detailed in the present paper including a short review of main characteristics of collisionless shocks (Sect. 2), the numerical constraints and initial conditions for simulating shock (Sect. 3), the methods for validating numerical simulation results of shocks (Sect. 4), the advantages of full particle simulations for analysing collisionless shocks (Sect. 5), and the analysis strategies used in the post processing stage (Sect. 6). In contrast, main results obtained with such simulations will not be presented herein but can be found in already published material (references are provided in the text). Some results will be only quoted herein (but not developed) when necessary, i.e. in order to illustrate a few aspects of these simulations.

Thanks to the increasing improvement of super computers, full particle simulations have proven to be quite helpful for analysing in detail the dynamics of collisionless shocks. The main characteristics of such codes can be shortly summarized as follows:

– One resolves the full set of Poisson and Maxwell's equations without any approximation. Two approaches are commonly used for resolving this equation's

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 54–78, 2003.

set, more precisely the space derivatives: *(i)* the finite difference approach and *(ii)* the use of FFT's (Fast Fourier Transform). Approach (ii) presents two advantages: first, FFT's are highly optimized in supercomputers libraries; second, FFT's allow to separate all fields components into two groups: the longitudinal electrostatic component E_l (solution of Poisson's equation) and the transverse electromagnetic components E_t and B_t solutions of Maxwell's equations (so called "fields pusher"). Such a separation is quite helpful in the post processing stage necessary for the data analysis, as will be outlined in this presentation.

– Both ions and electrons populations are treated as individual finite-size particles and suffer the effects of all fields via the Lorentz force, so called "particle pusher", which is applied to each particle. Because of the large number of particles commonly used, the particle pusher represents the most expensive part of the calculations on which most efforts of optimization needs to be performed (in terms of "vectorization" or of "parallelism"). Relativistic effects may be included in this force via the use of particle momentum. Each particle has three velocity components (v_x, v_y, v_z), but may have 1, 2 or 3 space coordinates (x, y, z) according to the dimension of the code of concern.

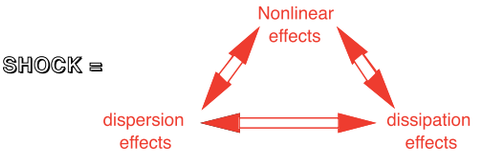
2 Main Features of Collisionless Shocks: A Short Review

For the purpose of clarity, we will remind shortly the basic features of a shock in order to establish more clearly the link between the internal physics of the shock ([39], [18]) and the specific numerical criteria necessary to shock simulations. More precisely, we will focus our attention on fast magnetosonic collisionless shocks. Corresponding full particle simulations of slow mode shocks will not be discussed herein, since these have been analysed until now with MHD and hybrid simulations mainly.

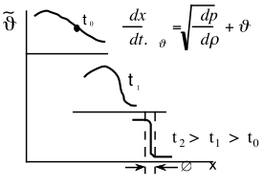
In a simplified and short approach, the shock can be expressed as the result of a compromise between nonlinear effects balanced both by dispersion and dissipative effects (Fig. 1).

2.1 Nonlinear Effects

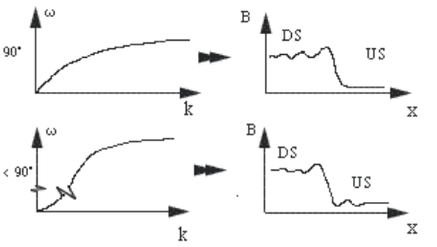
Assume a small amplitude sinusoidal wave with a given wavelength λ. In the linear regime, its velocity only depends on the wavefrequency ω and wavelength via the relationship $v = \omega/k$; the sinusoidal wave shape stays unchanged. As the amplitude increases, it reaches an amplitude threshold above which this relationship is not valid anymore. Then, the amplitude is so large that nonlinear effects become important and the velocity now depends on the local wave amplitude. In other words, one point of a high amplitude location will propagate with velocity larger than another fluid element with lower amplitude (Fig. 1a). As a result, the wave shape is now deformed with respect to the original sinusoidal pattern which is destroyed; this means that higher k-mode are excited (i.e. waves with smaller wavelengths are locally generated). This wave distortion is also called



a) Nonlinear effects : creation of smaller λ



b) Dispersion effects : elimination of smaller λ by radiation.



c) Dissipation effects: elimination of smaller λ by damping

Two types :	--> resistivity --> viscosity
Sources:	--> Macroscopic and microscopic processes

Fig. 1. Sketch illustrating the shock as the result of a compromise between three effects, namely the nonlinear effects (a) balanced by two others namely the dispersion effects (b) and the dissipation effects (c); "US" and "DS" define respectively the upstream and downstream regions.

"wave steepening" or wave "overtaking". This steepening means accessibility to smaller and smaller spatial scales. However, this steepening cannot continue forever; indeed, it reaches some scales over which other processes begin to be quite efficient and start to counterbalance this steepening. How is this balance established? Two complementary effects contribute.

2.2 Dispersion Effects

When higher k-modes waves are generated by the steepening (i.e. are locally accumulated and start forming a shock front-like pattern), dispersion effects

allow to radiate these waves from the front itself. Each wave propagates with its own velocity. In other words, the shock front is acting as an "antenna". Within a fluid approach, this radiation is supported by the Hall-current term. Figure 1b illustrates the correspondence between the dispersion curve properties and the shape of the steepened magnetosonic wave (i.e. shock front). For a strictly perpendicular magnetosonic wave, the dispersion curve is purely concave and higher k-modes do have lower phase and group velocities; the waves can only radiate downstream from the front. There is no upstream precursor. However, for an oblique propagating magnetosonic wave, the dispersion curve exhibits a convex shape from the origin (i.e. for lower k-modes) and a concave shape (for higher k-modes). The convex part will allow some waves to propagate with higher phase and group velocities and to escape from the front upstream. This will give birth to the upstream precursor. In contrast, the concave part will still contribute to the downstream evacuation.

2.3 Dissipation Effects

The analysis of these effects is the first difficult question in shocks physics since quite different source mechanisms can contribute; the difficulty consists in identifying these mechanisms. In short, two classes of dissipation processes may be defined: the resistive and the viscous effects. The relative contribution of these effects to the overall shock dynamics strongly varies according to the Mach regime, the propagation angle of the shock and the plasma beta number. Such mechanisms have been largely developed in previous review papers and text books and will not be detailed herein. However, it is necessary to focus on the basic processes contributing to these effects in order to identify which numerical criteria have to be satisfied in the simulations of collisionless shocks. We can identify two main particle interactions processes:

a) Macroscopic processes: these concern the interaction of particles with macroscopic E and B fields at the shock front. The origin of the electrostatic field E is due to the fact that ions and electrons do have quite different masses. In a simple approach, the electrons can be considered as strongly magnetized in contrast to ions which are partially unmagnetized when penetrating the shock front. As a consequence, some space charge effects (and associated electrostatic field E) build up at this front. The sign of the E field is appropriate for braking the ions and accelerating electrons through the front; in other words, the front plays the role of a "filter". For a strictly perpendicular front, all electrons are directly transmitted; in contrast, only ions which have enough energy succeed to pass through the front, while a certain percentage (depending on the Mach regime) is reflected. However, for an oblique propagating shock, the situation strongly differs since each population (electrons and ions) divides into two parts at the shock encounter: one part is reflected, the other is directly transmitted. The role of "filter" is drastically changed. In a deeper approach, the situation is more tricky and requires a more refined analysis which is out of scope of the present analysis.

b) Microscopic processes (Fig. 2): a certain number of instabilities may coexist at and in the near vicinity of the shock front. These represent sources of free energy sources; during their relaxation times, this free energy is redistributed in a self consistent way via waves-particles interactions. For the purpose of clarity, these instabilities can be gathered into three main groups:

– Current-driven instabilities. These instabilities may be due to field-aligned currents or cross-field currents. The main processes are the following. In short, the shock front corresponds to a large jump in magnetic field which is supported by an important diamagnetic cross-field current. This current mainly carried by electrons (which suffer different types of drifts) is large enough to trigger several kinds of instabilities (depending on propagation angle, local plasma parameter, etc..). The same holds for field-aligned currents. Then, electrons and ions are interacting with such instabilities which contribute to their local acceleration and heating (in addition to those brought by direct macroscopic processes). In summary, the current energy is dissipated by a transfer to particles via these instabilities; this represents the physical background of anomalous resistive processes. One advantage of full particle simulation is to include such processes by a self-consistent way as explained in Sect. 3.2.

– Anisotropy-driven instabilities. When electrons and/or ions interact with macroscopic fields (as mentioned above), these acquire different energy gains in directions perpendicular and parallel to the local magnetic field. This gives birth to noticeable temperature anisotropies which are large enough to trigger some instabilities. The particles interact back with these instabilities which relax in time and in space over different scales at and around the shock front as long as the anisotropy stays above a certain threshold. Then, it is important to prepare appropriate simulation conditions if one wishes to include by a self-consistent way such instabilities in the simulation.

– Non-equilibrium distribution instabilities. These instabilities represent a second consequence of the interaction of particles with macroscopic fields at the shock front. For ions, typical examples are: formation of ion ring distribution which characterizes the portion of incoming ions reflected at the front (quasi-perpendicular shock), or formation of field-aligned ion beams for ions which are reflected in quasi-parallel shocks (ion foreshock). For electrons, local depletions are common in electron distributions within the shock ramp, or loss-cone distributions (for low energetic reflected electrons) and field-aligned beam distributions (for high energetic reflected electrons) in the upstream electron foreshock. These instabilities also relax over different time/spatial scales and dissipate their energies by interacting back with both particles species.

Anisotropy-driven instabilities and non-equilibrium distributions instabilities are mainly based on velocities gradients processes and may be considered as sources of anomalous viscosity.

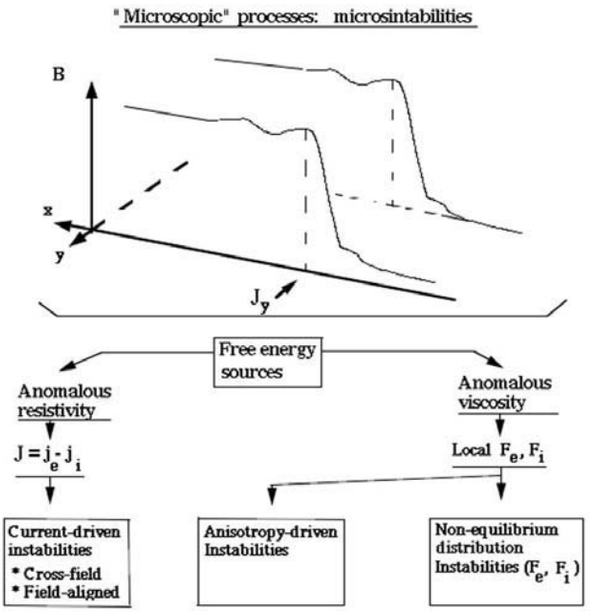


Fig. 2. Sketch illustrating the various sources-mechanisms responsible for microscopic processes.

3 Simulation Constraints for Simulating a Shock

When simulating a collisionless shock with a full particle code, different conditions need to be satisfied and are summarized as follows:

3.1 Common Criteria for the Code Stability

A certain number of criteria needs to be satisfied for the code stability itself and for describing space charge effects and magnetic effects on each particle motion with enough accuracy. These criteria are common to any particle code based on individual finite size particle; these can be found in many text books ([1],[3], [7], [17]), in previous ISSS proceedings ([9] to [15]) and in the review presented by Pritchett (2002, this issue). These will not be detailed herein.

3.2 Initial Conditions Inherent to Shock Features

Some additional criteria need to be verified in order to satisfy the time and spatial scales of the shock itself with appropriate accuracy and to include scales of specific instabilities inherent to the dynamics of the shock front. Taking into account the features of collisionless shocks summarized in Sect. 2, the main numerical criteria specific to shock dynamics can be summarized in two classes

corresponding to macroscopic and microscopic processes (upstream plasma parameters are used as reference values), respectively.

Numerical Criteria Related to Macroscopic Processes

a) Since the dynamics of the shock itself is strongly controlled by kinetic effects of the ions (in particular by ion gyration), the time length of the whole run must be much larger than one ion gyroperiod ($T_{run} >> \omega_{ci}^{-1}$, where ω_{ci} is the upstream ion gyrofrequency)

b) The size L_{plasma} of the plasma box needs to be much larger than spatial scale covered by the shock during its overall propagation with velocity V_{shock} within the plasma box ($L_{plasma} >> V_{shock}/T_{run}$).

c) Under certain conditions, some ions and/or electrons are reflected upstream by the shock front and acquire a velocity v_{part} much larger than the shock velocity V_{shock}. In addition, for oblique shocks, some wavetrain precursor may also be emitted from the shock ramp (Sect. 2) with a velocity V_{prec} also larger than V_{shock}. Then, the plasma box size needs to satisfy $L_{plasma} >> v_{max}/T_{run}$, where v_{max} denotes the largest value between v_{part} and V_{prec}.

d) Additional conditions need also to be satisfied by the length size L_{plasma}. First, space charge effects (via the build up of the electrostatic field) play at the shock front an important role on particles dynamics in association with the magnetic effects (Sect. 2); so do the inertia effects as well. As a consequence the size L_{plasma} must be much larger than the Debye length ($L_{plasma} >> \lambda_{De}$, λ_{Di}), the particles gyroradii ($L_{plasma} >> \rho_{ce}$, ρ_{ci}), and the inertia lengths ($L_{plasma} >> c/\omega_{pe}$, c/ω_{pi}). Plasma parameters must be chosen so that each characteristic plasma length is described with enough accuracy i.e. is larger than the unit grid size Δ.

Numerical Criteria Related to Microscopic Processes

As summarized in Fig. 2, three groups of microinstabilities intrinsic to the shock dynamics are expected. In a linear theoretical approach, each one is characterized by its own growth rate (γ) and wavelength (λ_{inst}). Such information may be accessed by solving the corresponding dispersion relation. At least, T_{run} and L_{plasma} values must be chosen so that the time length of the run satisfies $T_{run} >> \gamma^{-1}$, and the size of the plasma box in any direction verifies $L_{plasma,x,y} >> \lambda_{inst}$. Let us remind that such instability characteristics are defined within a linear theoretical approach and for homogeneous plasma conditions. These conditions are far from those intrinsic to the shock i.e. where nonlinear and inhomogeneous effects are present. However, such information are helpful in order to get some basic informations before starting the simulations.

Some simple approach may be also performed for identifying such instabilities by using the intrinsic features of these instabilities and/or simple geometric considerations. One example may be given for the cross-field current instabilities triggered at the shock front as shown by [4] and [21]. Simulating a strictly

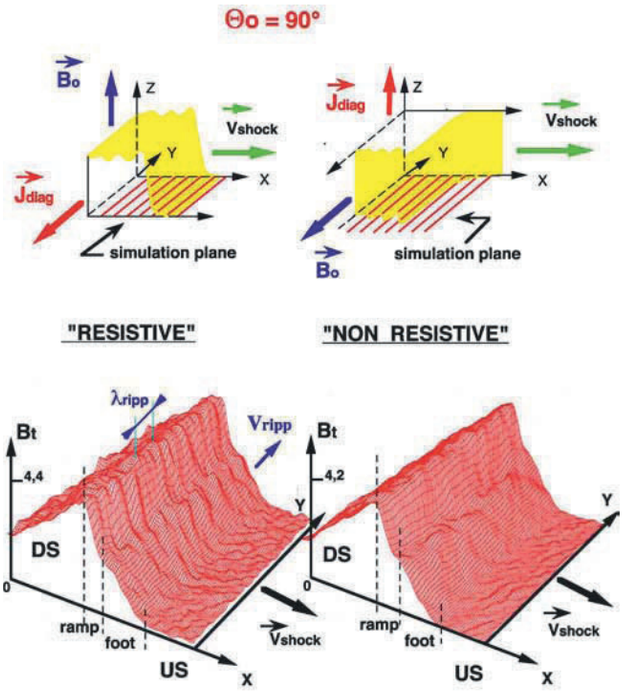


Fig. 3. Sketch illustrating the two different configurations of the magnetostatic field for simulating a strictly perpendicular shock when the upstream magnetostatic field B_o is outside (resistive case including the cross-field currents instabilities) and inside (non-resistive case) the simulation plane. A front rippling clearly appears characterized by a scale length λ_{ripp} in the resistive case.

perpendicular 2D (x,y) shock (propagating along x axis) may be performed with two different orientations of the magnetostatic field B_o (Fig. 3). First, the B_o field is contained within the 2D simulation plane (aligned along y axis); then, the cross field current is outside this plane (along z axis) and instability contribution is excluded from the simulation (non-resistive case). Second, the B_o field is outside the 2D simulation plane (along z); then, the cross-field current is inside this plane (along y) and instability contribution is self-consistently included in the simulation plane (resistive case). These instabilities are responsible for the front rippling. If the instability is still in the linear regime, this rippling may be characterized by a wavelength λ_{ripp} relatively easy to be measured; this rippling is often moving with a velocity which can also be measured. Both measurements allow to access the frequency ω_{ripp} characterizing this rippling, which can be compared with frequencies of the expected instabilities. By using this procedure, it was shown that the front rippling is related to the presence of lower hybrid frequency waves triggered by cross-field current instabilities [21]. However, for obliquely propagating shocks, the rippling may be the contribution

of different types of waves (as whistler precursor emitted from the ramp). In this case, a more sophisticated approach based on local waves analysis is necessary in order to identify the different waves components. In most cases, the existence or the propagation features of instabilities require more than one dimension in real space, i.e. 2D or 3D codes are necessary.

One difficulty is to include all possible instabilities for both electrons and ions. However, such a difficulty may also be used as an advantage for focusing on a given "dominant" instability selected by some appropriate way for instance. This "selection" can be performed by using a judicious choice of plasma parameters, of orientation of the magnetostatic field, and of the growth rate and wavelength of the concerned instability (estimated by resolving the corresponding linear dispersion relation). The growth rate can largely vary between different instabilities; the time length of the total run and the size L_{plasma} represent other parameters to play with, for including or excluding a given instability. As a consequence, the impact of the "selected" instability may be analysed on both electrons and ions by a self-consistent way.

In summary, due to some computer constraints, some compromise is necessary for including most macroscopic and microscopic processes, and may be obtained by using nonrealistic values of mass ratio or of ratio ω_{pe}/ω_{ce}.

3.3 Initial Conditions for Exciting a Shock

Different initial conditions can be used for exciting a collisionless shock. These conditions often correspond to the first delicate phase of the simulation. Up to now, five different methods can be identified (Fig. 4):

Relaxation Method (with Forced R-H Conditions)

This method has been initially proposed by [26], [27] in order to simulate hybrid simulations of a shock. The initial state consists of two uniform regions separated by a thin intermediate layer. In the upstream region (LHS for instance), ions are uniformly distributed along the real axis x and given random velocities to approximate a Maxwellian distribution convecting toward the intermediate region with appropriate density (n_1), temperature (T_{i1}) and flow speed (V_1) along x. The upstream magnetic field (B_1) and electrons temperature (T_{e1}) are also assumed to be uniform. The downstream region is prepared similarly with density (n_2), ion temperature (T_{i2}), flow speed (V_2), magnetic field (B_2) and electron temperature (T_{e2}). The thin intermediate region is prepared such that the density is a linear function of x which matches n_1 at the left and n_2 at the right end and satisfies $B/B_1 = n/n_1 = V_1/V_x$, some temperature profile being assumed. The downstream quantities are computed from the upstream values using the Rankine-Hugoniot relations with γ=5/3 and an assumed initial value of T_{e2}/T_{i2} [39]. The aim of this method is to prepare at time t=0 a shock transition sufficiently close to the final state so that the system can eventually reach its final state by allowing the shock transition to evolve in time. The initial downstream state is by necessity an approximation. Since this pioneering

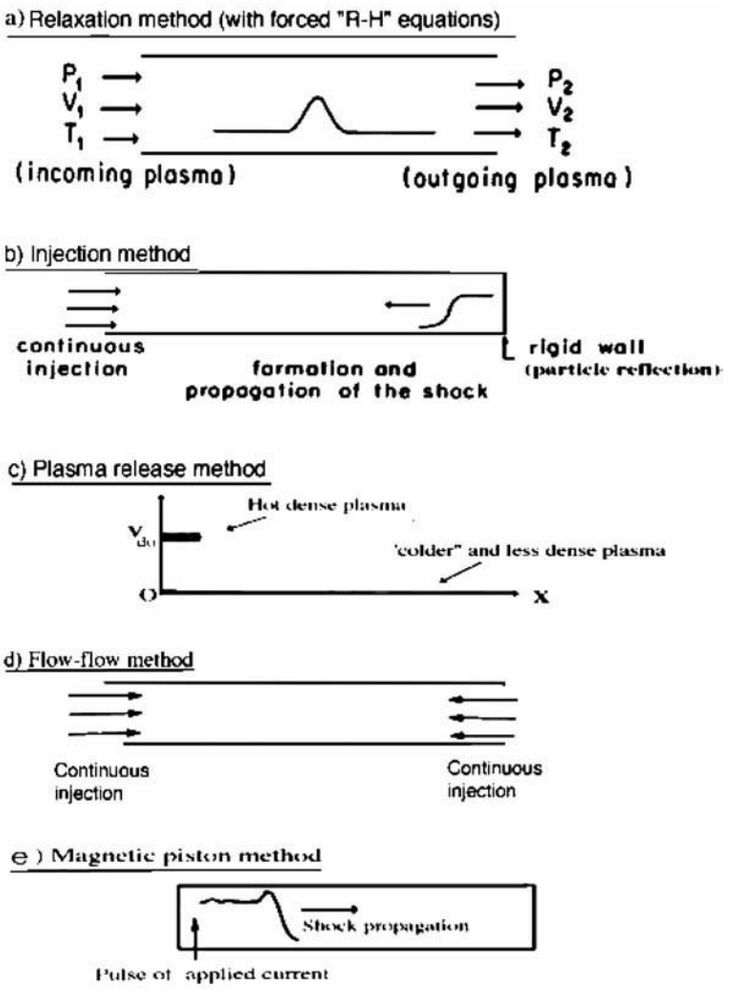


Fig. 4. Summary illustrating the five different simulation methods used for exciting a collisionless shock.

work, this method including several constraints has been more or less abandoned, and has been replaced instead by others listed below. One reason could be the fact that the R-H conditions are initially forced in the system instead of being reached by a self-consistent way after a certain time; this time must be large enough so that the formed shock is independent of initial conditions. Without such constraints, other methods are quite appropriate to analyse the nonstationarity of the shock around and/or far from the front for instance. Another reason could be the relative simplicity of the other methods used nowadays.

Injection Method

A plasma is injected from one side (LHS for instance) of the simulation box and is reflected back when it reaches the other side (RHS), where specular reflection conditions are locally applied (Fig. 4b). The reflected plasma interacts with the incoming plasma and an interface forms between both plasmas. This interface allows the jump transition in all local plasma momenta and in field components between upstream and downstream regions. As time evolves, this interface transforms into a shock which propagates from RHS to LHS [8], [38].

Plasma Release Method

This approach includes initially two different plasmas. It consists releasing a hot and dense plasma (animated with an initial bulk velocity at time t=0) into a relatively colder and less dense ambient plasma (at rest) as illustrated in Fig. 4c. At the frontier between both plasmas, an interface builds up accompanied by a diamagnetic cavity which extends behind the propagating hot plasma component. As time evolves, this interface propagates from LHS to RHS, transforms into a collisionless shock, and reaches a regime independent of initial conditions. This method has been adopted with 1-D full particle code by [29], [30], [31] and more recently by [23], and with 2-D full particle code by [16]. It reveals to be quite adapted for analysing shock dynamics and/or interface instabilities in situations including plasma releases in solar physics such as CME (coronal mass ejections), in active spatial experiments (AMPTE and CRESS missions) and also in laboratory fusion experiments in which a dense and hot plasma (resulting from a laser ablated solid target) is expanding within an ambient plasma. One advantage of the method is that the initial bulk velocity of the hot plasma component is providing directly an estimate of the Mach regime for the shock. In many cases, the abrupt transition between both plasmas is fixed at time t=0 over one grid point only. However, it is recommended to smooth out slightly this density gradient over several grid points in order to get a more progressive transition between both plasmas.

Flow-Flow Method

Another method slightly different from injection and plasma release is the so-called "flow-flow" method in which two counterstreaming plasmas are launched each occupying the half of the box [32]. The plasma is continuously injected from both sides of the simulation box with a given bulk velocity. After some time, the two plasmas begin to couple and form a pair of interfaces. A pair of shocks is formed at later times, which propagate in opposite directions away from the center of the box.

Magnetic Piston Method

In this case, an external current pulse is applied and added to the self-consistent plasma current in Maxwell's equations, which induces a pulse in electric and magnetic fields components. This method has been intensively applied for 1D [19], [20] and 2D full particle simulations [21], [34] and references therein. The pulse is propagating into the plasma and develops into a shock. As time evolves, the shock front separates shortly from the propagating piston itself. This method has several advantages: it is simple (only the location, the amplitude and the characteristic time length T_{app} of the pulse need to be defined as initial conditions), and rapid (the shock builds up in a short time and becomes independent on initial conditions over a time range much smaller than one upstream ion gyroperiod). A few more practical precisions concerning the application of the current pulse need to be provided as follows (Fig. 5):

a) The whole simulation box is divided into two parts: one is dedicated to vacuum (no plasma), the other includes the plasma, as illustrated in Fig. 5a. The pulse is applied within the vacuum region over a few grid points located close the left hand side of the plasma region. As the pulse is applied, two pulse-shaped waves are generated which propagate in opposite directions. One pulse (called "compression" wave) enters almost immediately the plasma box, acts as a "piston" and generates a shock front in front of it; the other pulse (called "dilatation" wave) is propagating within the vacuum region. Some appropriate boundary conditions – described in Sect. 3.5 – need to be included in order to "kill" this undesirable wave.

b) the second point concerns the time range over which the current pulse is applied. In practise, it is important to apply this pulse smoothly in time so that the plasma has enough time to "admit" this perturbation without generating some spurious "unphysical" or "undesirable physical" signatures. In practise, the pulse amplitude is varying in time as follows: the amplitude varies as a half-sinusoidal function within the time range $(0, T_{app})$, and is constant for the (T_{app}, T_{run}). Time T_{run} is the time length of the whole simulation run, while T_{app} is a time value which needs to be adjusted (in practise, several hundreds of time steps are recommended). One easy way to test whether T_{app} value is appropriate consists in simulating a strictly perpendicular shock ($\Theta_o = 90°$) in the supercritical regime (with a reasonable simulation box size). If T_{app} is too short, a precursor is emitted from the ramp (T_{app}=5 in case 1 of Fig. 5b) propagating at the light velocity. This is in contrast with the fact that, for $\Theta_o = 90°$, no upstream precursor is expected (Sect. 2). In fact, this precursor has a physical meaning. When the pulse application is too rapid, high-k modes are generated; as a consequence the second term (k^2c^2) of the wave dispersion relation $\omega^2 = \omega_{pe}^2 + k^2c^2$ corresponding to the light wave is not negligible any more and is supporting this light emission. One simple way to avoid this emission consists in increasing T_{app} (=30 for instance) in order to smooth out the pulse application, as illustrated in case 2 of Fig. 5b.

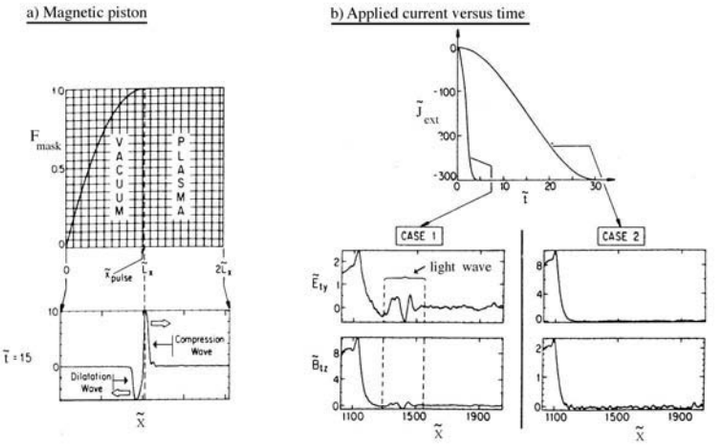


Fig. 5. Sketch illustrating some details of the magnetic piston method. (a): configuration of the simulation box into two parts including respectively a vacuum and a plasma part; the absorbing function F_{mask} is equal to 1 in the plasma part and decreases from 1 to zero value according to a quadratic law in the vacuum part. (b): impact of the applied external current pulse on the shape of transverse electric and magnetic fields, when the pulse application time is short (case 1: $0 < t < 5$) and larger (case 2: $0 < t < 30$).

As a consequence, it is recommended not to apply a time varying pulse which corresponds to a too sudden perturbation. Let us precise that the shock front itself is characterized by a large jump in magnetic and electric fields components in association with large variations in plasma momenta from low (upstream) to high (downstream) values. The piston itself is characterized by a very rapid variation of the plasma density from high to almost zero value when moving further into the downstream region; using a simple picture, it acts a "snow-plow" machine

Each method has advantages or inconveniences in terms of relative simplicity and how quick the shock is formed. In any case, the main purpose must be satisfied: same characteristics of the shock need to recovered (for similar Mach regime and plasma conditions) for any used method, as soon as the shock dynamics becomes independent on initial conditions.

3.4 Full Particle Simulations of Planar and Curved Shocks

All above methods of Sect. 3.3 have been mainly applied for initiating a planar-like shock in 1-D (with infinite shock width) and 2-D (with finite shock width) simulations. Similar technics can be also applied to self-consistent curved shocks. However, the number of curved shock simulations is quite limited and restricted to a few works only. One is based on 2D hybrid simulations ([40]; Winske and Omidi, this issue), where the shock is initiated by using the reflecting-wall type method: the flat reflecting obstacle (i.e. the reflecting wall) has been replaced by

a curved obstacle so that an expanding curved shock wave is formed after the mutual interaction of incoming and reflected plasmas takes place. The other is a 2-D full particle simulation of curved shocks performed by [36], [37], based on magnetic piston method. In this last example, the pulse current is now applied over a few grid points distributed within a small cylinder, in order to initiate the curved shape of the shock. The critical question concerns the boundary conditions: indeed, due to the constraints of up-to-date computer facilities, these full particle simulations have been limited at present time to quasi-perpendicular curved shocks (where $90° < \Theta_o < 45°$). Simulations including the full angular range $90° < \Theta_o < 0°$, i.e. when the full magnetostatic field is within the 2-D simulation plane, are quite difficult to be performed at the present time. Indeed, more and more particles are reflected by the shock front for strongly oblique directions and are streaming away along the magnetic field into the upstream region (self-consistent formation of the foreshock). Then, high energetic reflected particles reach within a short time the opposite side of the simulation plasma box, and the simulation has to be stopped. This problem can be resolved by a "trick" which consists in following the motion of particles within the real space projected into the 2-D simulation plane, instead on following the same full motion within this plane itself. One consequence is that the field-aligned motion is intact (in terms of velocity dynamics since each particle always has three individual velocity components), but will be extended over a reduced distance in real space, which is in favour for using a large but reasonable simulation box size. In other words, one performs "pseudo 3-D" full particle simulations. For so doing, the upstream magnetostatic field is characterized by two Euler angles: one angle corresponds to the tilt-angle within the plane (necessary for oblicity), the other corresponds to a tilt-angle out of this plane. The combination of both angles allows to control easily the extend of the curved shock and of the resulting foreshock. Such method has allowed to reproduce and to analyse in detail the electron foreshock [37]. A refinement of this method has recently allowed to extend the restricted range to $90° < \Theta_o < 30°$ in order to analyse self-consistently the transition from quasi-perpendicular to quasi-parallel angular domains.

3.5 Boundary Conditions Specific to Shocks Simulations

Very often, some specific boundary conditions need to be adapted since a shock corresponds to a jump of entropy; the origin of the entropy production by turbulence is the time behaviour of the distribution function [33]. Indeed, characteristics of plasma in upstream and downstream regions are quite different and due to irreversible processes. Then, two types of boundary conditions are necessary: one for fields components, the other for the particles dynamics. These boundary conditions represent the second delicate phase of the simulation:

a) Field components: full particle simulation codes may be based on finite different technics or FFT's technics (Fast Fourier Transform), in order to resolve the space derivatives involved mainly in Poisson-Maxwell's equations.

One can use FFT's even for collisionless shocks simulations where two different states of the plasma are fully involved in the plasma box (in upstream and downstream regions respectively). However, one has to suppress artificially any possibility for downstream ingredients to be injected in the upstream region or vice versa because of the intrinsic periodicity features. One example can be given for the magnetic piston method presented in Sect. 3.3. As illustrated in Fig. 5a, the dilatation pulse-like wave propagates within the vacuum box but, when reaching the LHS wall at (x=0), may reenter the plasma box at its RHS (at x=2 L_x). Then, this dilatation pulse may perturb the upstream region of the shock, which is not desirable. One way to resolve this problem consists in applying an artificial absorbing or "masking" function F_{mask} everywhere in the simulation box: F_{mask}=1 within the whole plasma box so that all fields components are unchanged. However, F_{mask} varies according to a quadratic function from 1 to zero (or small value) when moving from the RHS to LHS of the vacuum box (Fig. 5a).

How to apply this function? Experience shows that a few fields components only, -and not all fields-, need to be multiplied by this function. Indeed, let us stress that it is important to leave the plasma with some degrees of freedom i.e. to let the self-consistent effects act. In summary, the method consists in "killing" a few fields components in such a way that other components will be "killed" self-consistently.

b) Particle dynamics: let us consider again the magnetic piston method and Fig. 5 as a reference. Within the plasma box, downstream and upstream regions are extended respectively from the LHS wall to the shock front, and from the shock front to the RHS wall. During their thermal motion (plasma at rest) some upstream particles may reach the RHS wall, and need to be treated appropriately in order to avoid any artificial particles accumulation there (spurious local space charge effects), or any reinjection back into the vacuum region (via x=0) in the case some periodic conditions are used in the code. One simple way to avoid these problems consists in applying specular reflection conditions for any particle reaching this RHS wall (at x=2 L_x). On the statistical viewpoint, experiences show the number of particles which suffer such reflection from v_x to -v_x (in 1D code for instance) is balanced by particles being reflected from -v_x to v_x so that no local accumulation can build up. This point can be easily checked by plotting an enlarged view of local particles density or of phase space plots around the RHS wall.

The situation is more simple on the LHS wall of the plasma region (at $x = L_x$ in Fig. 5a). Indeed, as the piston acts as a "snow-plow machine" during its propagation, the plasma density behind it is almost zero. However, some particles may be still present behind, and for safety same specular reflection conditions are also applied on particles which could reach the LHS wall. Experiences show that there is no risk for these "artificially" reflected particles to reach back the downstream region of the shock nearby the front, since the shock front is propagating at velocity much faster than the piston itself. However, it is recommended

to check this point (with enlarged plots of phase space around the wall) for a new simulation practitioner.

4 Methods for Validating Numerical Simulation Results on Collisionless Shocks

In practise, applying all conditions of Sect. 3 leads to find a compromise in the choice of plasma parameters. As already mentioned in Sect. 3, non-realistic values of the mass ratio and the ratio ω_{pe}/ω_{ce} of electron plasma frequency over gyrofrequency are commonly used. In order to validate the results issued from this compromise, different strategies can be followed:

– Validation by performing simulations runs for similar shock and plasma conditions with full particle codes of different dimensions (1D or 2D for instance, as commonly used at present time). This method may be applied to confirm any process which is independent of shock dimension (cyclic self-reformation of the shock front for instance; [2], [19], [21]. Later on, this type of validation has the advantage for performing many simulation runs at a lower cost i.e. by using the lowest dimensional code.

– Validation by performing a parametric study on some "sensitive" plasma parameters in order to recover a similar process over the corresponding scale. For instance, the cyclic period of the shock front self-reformation mentioned above is always of the order of the mean ion gyroperiod measured in the ramp (for a quasi-perpendicular shock in super-critical regime). This feature and associated relative scales can be easily recovered by performing various full particle simulations based on different mass ratios, since the ion gyromotion (which drives the self reformation process) is very sensitive to the mass ratio. Same procedure can apply for confirming the size of the trapping loop of reflected ions (of the order of the local ion gyroradius) in a supercritical shock regime.

– Validation by comparing results issued from hybrid and full particle simulations. This comparison is a delicate question which has only been approached recently [23]. One of the main results is that similar shock dynamics (except electron kinetic dynamics) can be recovered in both types of simulations provided that some appropriate readjustment on the grid resolution size and on some plasma parameters is performed.

– Validation by developing a simple analytical model dedicated to a particular process. For instance, a simple hybrid-type model has been developed recently and allows to show that the nonstationary behaviour of the shock front (due to its cyclic self-reformation) depends mainly on the density of reflected ions accumulated upstream. In other words, this mechanism is inherent to the shock front itself, and is independent on the mass ratio and on ω_{pe}/ω_{ce} ratio used in full particle simulations [6].

Moreover, one basic and additional validation of the results consists in verifying the conservation of total energy versus time (fields and particles) in the system.

5 Advantages of Full Particle Simulations for Analysing Collisionless Shocks

A key advantage of full particle simulations for analysing collisionless shocks is that all physical scales (both ions and electrons) are fully accessible. This accessibility appears in different features as illustrated below:

5.1 Evidence of the Shock Front Turbulence

The shock front exhibits a strong nonstationarity which represents one typical problem to be analysed in details with the help of the multi-satellites mission CLUSTER-2, as illustrated in a computer-generated movie shown during the ISSS-6 meeting. As an example, two different time/spatial scales have been identified in full particle simulation results: one (ion scale) is due to the cyclic self-reformation of the shock front, the other much smaller than the ion scale (not clearly identified yet for oblique propagating shocks) is due to the propagation of the front rippling along the shock front direction [4], [21].

5.2 Evidence of Upstream Precursor

In full particle simulations, this precursor may easily access to smallest and largest physical scales such as electron scales (for oblique quasi-perpendicular shocks) or ion scales (for oblique quasi-parallel shocks). Analysing the impact of each population on this precursor (and of associated wave steepening and damping) is quite possible since the dynamics of both populations is fully involved. The precise identification of waves contributing to this precursor in different Mach regimes is under active investigation at the present time, in order to be compared with experimental data of the multi-satellites mission CLUSTER-2.

5.3 Appropriate Scaling of the Shock Ramp

The ramp thickness is shown to be larger than electron inertia length but smaller than ion inertia length. Accessibility to such scales via magnetic and electric field profiles do have an impact on the acceleration and heating of particles passing through the front [22]. A similar feature is expected for reflected particles too. One key point is that this thickness varies self-consistently in time according to the different turbulence stages of the front (for a fixed Mach regime); in addition, it adjusts itself to different plasma conditions and propagation directions of the front.

5.4 Appropriate Description of the Full Electrons Dynamics

This property is inherent to the fact that all kinetic effects are fully included. Full particle simulations reveal to be an essential tool to analyse the electron dynamics via a self-consistent approach. Thanks to the increasing power of super-

computers facilities, this analysis has been possible. As recent typical examples, these simulations have allowed to analyse in great detail (i) the source-mechanisms responsible for the breakdown of adiabaticity of transmitted electrons [25], (ii) the reflection of electrons at the front [24], (iii) the electron pre-heating in the whistler precursor of oblique shocks for various Mach regimes [28], [35], or (iv) the formation of the electron foreshock in a curved shock simulation [37]. A few examples will be illustrated in Sect. 6.

6 Strategies of Analysis Used in Post-processing Stage

Different types of analysis may be developed in the post-processing stage. We will distinguish two approaches which require complementary post-processing: analysis of the basic mechanisms involved in both particle and field dynamics, and analysis dedicated to "mimicing" measurements performed on board of a real space mission.

6.1 Strategies Used for Analysing Basic Mechanisms from Simulation Results of Collisionless Shocks

Results issued from full particle simulations of collisionless shocks present sometimes complex structures and/or phenomena which require different but complementary approaches. At the present time, four different approaches can be identified:

– Comparison of results obtained from runs using different dimensional codes or magnetic field configurations. This allows to include or exclude in the simulations some effects at will. A first example is the study of the source mechanism responsible for the shock front rippling where the orientation of the static magnetic field is in (no rippling) or out (rippling) the simulation plane as illustrated in Fig. 3. A second example is the impact of the shock front rippling on electrons dynamics analysed with $2-D$ code (inhomogeneous shock front) and $1-D$ code (homogeneous shock front), where $1-D$ and $2-D$ simulations are performed for similar shock conditions and with the same plasma parameters [24]. Such a procedure allows to analyse and to quantity the impact of the shock front rippling in particles dynamics. A third example is the comparison of $2-D$ results between self-consistent planar and curved shock which allows to analyse the impact of curvature on particle dynamics [36], [37].

– Preselected particles trajectory (or "PPT") analysis. This method reveals to be quite helpful for analysing processes of high complexity by using a large number of diagnostics. It has been recently used for analysing the reflection mechanisms of electrons encountering a shock front [24]. It requires two successive runs and allows to follow particles trajectories without any a priori assumption. In the first run, reflected electrons are identified among all electrons and "marked". A second run identical to the first one is performed during which most information of preselected (i.e. previously "marked") electrons are stored as a function of time. These information include the usual quantities as locations

and velocities components, but also the fields and field gradients components experienced by each "marked" particle. These last information allow to analyse the basic drift mechanisms that each individual particle is suffering versus time and versus its relative location within the shock front.

This method is fully self-consistent (in contrast to a "test particles" approach) and presents two advantages: first, it allows to perform an "individual particle" approach in order to identify precisely some typical trajectory within the shock front and the underlying acceleration mechanism. By this way, three main classes of mechanism responsible for electrons reflection have been identified: magnetic mirror (or fast Fermi) process, trapping process and multi-bounces process (which combines the two first mechanisms). Second, it allows to perform a "statistical" approach in order to identify some particular groups of electrons. This approach has demonstrated the formation of electron bursts in time, and of electron packs in space (i.e. reflected electrons form beams which are not homogeneous in space). As a consequence, a field-aligned electron beam resulting from electron reflection is not spatially homogeneous. This result is in contrast with initial conditions of most previous full particle simulations of wave emissions triggered by beam-plasma interactions.

– Classical "test particles" analysis. In some cases, the self-consistent results include several intricated mechanisms over quite different scales and reveal to be quite complex to analyse. Then, a complementary approach consists in "freezing" artificially a part of this self-consistency in order to recover a partial simplicity necessary for a better understanding of a given process. As an example, this approach has been quite helpful to identify the adiabatic / non-adiabatic processes applied to transmitted electrons passing through the shock front [25]. The method is using profiles of electric and magnetic fields not deduced from an analytical model but rather obtained from full particle simulations at a given time. Then, all electrons and ions spatial scales are already included in an appropriate way. A set of particles is launched against the shock front. Since the characteristics of these particles are under the control of the user, one can choose only particles in the core (low energy) or in the tail (high energy) of the velocity distribution at will in the initial conditions. Changes in trajectories and in velocities components are analysed individually and/or statistically, during and after their interactions with the shock front.

A judicious choice of field components profiles obtained from simulations may offer a more profound analysis of the underlying mechanisms of particle acceleration and heating at a relatively low computer cost. This is illustrated by considering three successive steps as follows. In a first step, let us consider a set of fields obtained from a 2-D full particle simulation where the shock front is propagating along the x axis, and is planar and nonhomogeneous (rippling) along the y axis. In a second step, let us consider all 2-D field components at a given time, and in a third step these same field components which have been y-averaged; in this last step, the front can be considered as one-dimensional, homogeneous and infinite along the y-axis. The third step allows to analyse the basic interaction mechanism of particles with their characteristic scales (gyro-

radii, inertia lengths, Debye lengths) versus the characteristic scales of the front (of precursor if any, of the foot, of the ramp and of the overshoot) with a simplified but still self-consistent shock. Comparison of results obtained by using field components obtained from steps 3 and 2 allows to quantity the impact of the front inhomogeneity at a fixed time. At last, the comparison of results obtained by using fields components obtained from steps 1 and 2 will provide information on the impact of the time-dependant field dynamics i.e. of the shock front turbulence. This strategy is presently under active use for analysing the dynamics of particles reflected by and transmitted through a shock front. It may be also applied for any other analysis.

– "Test particles" and "PPT" combined analysis. A further additional and complementary approach consists in analysing results issued from the above "PPT" method (self-consistent effects included) and the test particle method (self-consistent effects excluded). This combined approach can be also appropriate to determine time/spatial scales over which self-consistent effects play a key role or can be neglected. These information can be quite helpful before initiating complementary theoretical analysis where inclusion of self-consistent effects represents a great challenge and is often neglected.

6.2 Strategies Using "Virtual" Satellites "Mimicing" the Real Experiments

Once a simulation is performed, another question appears: is it possible to insert, in this simulation, the experimental conditions of local measurements specific to a given space mission? Of course, this question is not dedicated to full particle simulations only and can be extended to any type of simulation. If this approach could be systematically resolved, it would allow to establish a more direct link between the simulation and experimental results. Such a procedure has been initially developed for "mimicing" the crossing of a quasiparallel shock issued from 1D hybrid simulation code [5]. At that time, the shock crossing was restricted along the shock normal only, because of the use of 1D simulation results. Recently, some efforts have been invested in order to "mimic" the local measurements made by the four-satellites Cluster-2 mission crossing a 2D shock issued from 2-D full particle simulations. In this case, any direction of the shock crossing with respect to the shock normal may be mimiced. We will give two typical examples:

(i) in the first example, the purpose was to explain the different magnetic field signatures of the measurements made by each satellite during the shock crossing. This procedure (developed with AVS interactive graphic package) is summarized as follows. First, a 2-D full particle simulation of a supercritical shock has been performed. Nonstationarity of the shock front (Sect. 5.1) and corresponding spatial/time scales have been retrieved. The attention is focussed on the longest characteristic time scale; for doing so, we note the start and end times of a given cyclic self-reformation. Second, another run is performed identical to the first one, but where field components are stored at each (or a few) time step(s) between the start and the end time of the whole selected reformation

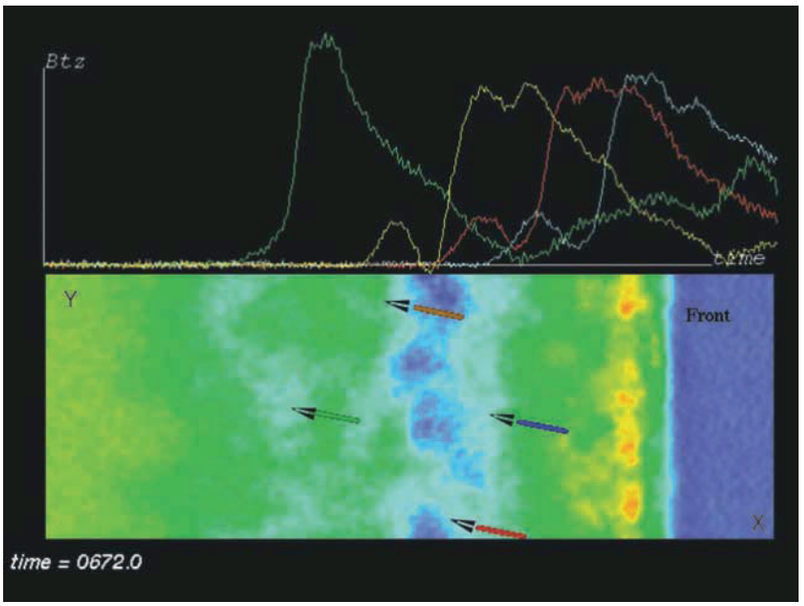


Fig. 6. (i) Bottom panel shows the view (from above) of the main magnetic field component $B_{tz}(x, y)$ and its crossing by four "virtual" satellites "mimicing" the Cluster-2 mission, at the last time of the animation (end of a cyclic self-reformation of the shock front). During the shock propagation from the left to the right hand side along x-axis, each satellite has moved over a distance indicated by a short straight coloured line with arrow. (ii) Top panel shows the local measurement of the $B_{tz}(x, y)$ magnetic component made by each satellite during the shock front crossing; the colour of each curve corresponds to that of the satellite tracking shown in bottom panel. Four different signatures of the $B_{tz}(x, y)$ component illustrate the fields turbulence at the shock front during its crossing. Satellites are at relative mutual distances of 600 Km approximately.

cycle. Let us focuss now on the main magnetic field component. Then, we have a first x-y plane containing these B(x,y) data with a high time accuracy; the size of this table corresponds to the size of the plasma box. Third, we separately develop an AVS tool defining a second x-y plane (same size as the plasma plane) including four points which can be driven independently from each other. Each point represents a "virtual" satellite and has, within this plane, dedicated starting location (x,y) and motion driven by velocity components, so that any type of shock crossing (inbound, outbound, oblique, normal crossing, etc..) may be reproduced independently of the shock motion. Fourth, a third plane is defined which includes the combined projection of both previous planes. Measurements of local B(x,y) field performed by four "virtual" satellites during a typical shock crossing are shown within this plane in Fig. 6. Four different signatures versus time of the main magnetic field component are clearly evidenced by the different satellites; these illustrate quite well the field turbulence which takes place

over different times and spatial scales at the shock front. Complementary waves analysis are now under active investigation in order to identify these waves.

How to optimize this post-processing techniques? This "mimicing" procedure offers different advantages and can be used in two different ways. The first one consists in reproducing experimental results with a given configuration defined by the real satellites. In other words, one only inserts within the simulation results the configuration approaching as much as possible the real conditions; the initial priority is motivated by data obtained by conditions of the real space mission. The second way consists in inverting this priority. As an example, let us focus on the small scale turbulence along the shock front. Its analysis requires to fix at least two "virtual" satellites aligned along the direction parallel to the shock front and distant each other over a scale L_{sat} larger than the characteristic scale λ_{ripp} (Sect. 3.2). Then, these satellites should cross the shock front almost simultaneously so that any waves activity along x-direction should be minimized in the signatures of the shock crossing. By adjusting the scale L_{sat} with respect to λ_{ripp}, one can analyse different types of waves contributing to the front nonhomogeneity; indeed, λ_{ripp} may vary according the Mach regime, the shock angle and the upstream plasma β conditions. This allows to define the best "satellites" configurations for analysing or identifying a certain type of wave components contributing to this local turbulence. Further, the next step will consist in searching among the whole set of shock crossing by CLUSTER-2, the configurations approaching that of the "virtual" satellites (for similar Mach regime conditions), and in comparing experimental and numerical signatures. In this case, the initial priority is motivated by results obtained by the "virtual" satellites. Similar procedures can be also used for local measurements of particles distribution functions; for so doing, 2D full particle simulations have been made recently based on a high number of particles per grid, in order to obtain appropriate local statistics.

(ii) the second example concerns the detailed analysis of the electron foreshock. In this case, any experimental measurement of the particle distribution function is "local" in space but involves particles originating from quite different regions, and which are counted "locally" at the time and location of the measurement. These electrons have been reflected from different locations of the curved shock front; different reflection processes are involved which can be identified with different signatures in the electron distribution. In recent 2-D full particle simulations of electron foreshock [37], such "time-of-flight effects" which play an important role have been self-consistently included. How to "mimic" the real measurement performed by the satellite? For so doing, a graphical procedure has been developed which consists by superimposing a curved grid adapted to the curved bow shock and aligned to the upstream edge of the electron foreshock. Such a procedure allows to keep the size of each grid constant when moving in angle or with upstream distance from the front; this feature cannot be obtained with the use of a radial grid set (centered at the shock curvature point). The size of the grid must be chosen in appropriate way (large enough to have credible statistics within each grid, and small enough to scan the whole foreshock region

with a minimum accuracy in angle and in distance from the front. Then, within each grid, local distribution is calculated and energy spectrum analysis is also performed. The next step (not made yet) consists in superimposing the trajectory of a "virtual satellite" to this curved grid set, "mimicing" the real satellite trajectory. This trajectory will select the set of successive grids within which the local electron distribution is measured; then, a direct comparison between experimental and numerical results will be possible. A refinement can be later obtained by adjusting the size of the space grid to the integration time of the real experiment (performed often over the satellite spin time period). Indeed, this feature has a full meaning since the size of each grid corresponding to the distance covered by the satellite during its full spin period is constant, whatever the angle along and the distance from the curved shock front are.

These examples illustrate the increasing importance of the post-processing stage and the efforts which need to be invested. More efforts are necessary at a time the number of particles involved in simulations and the size of associated field components tables are drastically increasing. In addition, multi-dimensional full particle codes will be also more commonly used with the increasing power of supercomputers and appropriate optimization of simulations codes. This means that improved graphics techniques (interactive 2D/3D if possible) with sophisticated and real time animations facilities (as virtual reality techniques) need to be encouraged and adapted to space plasma physics problems in order to analyse the huge amounts of stored data obtained from these large scale simulations.

References

1. Birdsall C.K. and A.B. Langdon, Plasma physics via computer simulation, McGraw-Hill, Singapore, 1985.
2. Biskamp D. and H. Welter, Ion heating in high-mach number, oblique, collisionless shock waves, Phys. Rev. Lett., 28, 410–413, 1972.
3. Dawson J. M., Particle simulations of Plasmas, Rev. Mod. Phys., 55, 403 (1983)
4. Forslund D.W., K. B. Quest, J.U. Brackbill and K. Lee, Collisionless dissipation in quasi-perpendicular shocks, J. Geophys. Res., 89, A4, 2142, 1984.
5. Giacalone J., S.J. Schwartz and D. Burgess, Artificial spacecraft in hybrid simulations of the quasi-parallel Earth's bow shock: analysis of time series versus spatial profiles and a separation strategy for Cluster, Ann. Geophys., 12, 591, 1994.
6. Hada T., M. Onishi, B. Lembege and P. Savoini, Shock front nonstationarity of supercritical perpendicular shocks, J. Geophys. Res., submitted to, 2002.
7. Hockney R.W. and J.W. Eastwood, Computer simulation using particles, IOP Publishing Ltd, New York, 1988.
8. Hoshino M. and B. Lembege, Formation of non-thermal, high energy particles in relativistic magnetosonic shock waves, Adv. Space Rev., 15, No 8/9, 67-70, (Proceedings of the D2.1 Symposium of COSPAR Scientific commission D, 30th COSPAR meeting, Germany), 1995.
9. ISSS-1, Computer Simulation of Space Plasma, Proceedings of the First International School for Space Simulation, Ed. H. Matsumoto and T. Sato, D. Reidel Publishing Company, Dordrecht/Boston, 1985.

10. ISSS-2, Space Plasma Simulations, Proceedings of the Second International School for Space Simulation, ed. M. Ashour-Abdalla and D.A. Dutton, Reidel Publishing Company Dordrecht/Boston, 1985; (reprinted from Space Science Rev., Vol. 42, Nos 1-4, 1985).
11. ISSS-3, Numerical Simulations of Space Plasmas, Proceedings of the Third International School for Space Simulation: Tutorial Courses, Eds. B. Lembege and J. Eastwood, North-Holland, Amsterdam, 1988; (Reprinted from computer Physics Communications, Vol. 49, N°1, 1988).
12. ISSS-3, Tutorial courses, Proceedings of the Third International School for Space Simulation: Tutorial Courses, Part1, Cepadues-Edition, Toulouse, France, 1989.
13. ISSS-4, The 4th International School for space simulation, Proceedings of the 2nd week symposium, March 25-April 6, Kyoto and Nara, Japan, 1991.
14. ISSS-4, Computer Space Plasma Physics: Simulation Techniques and Software, Terra Scientific Publishing Company, Tokyo, 1993.
15. ISSS-5, Proceedings of the 5th International School/Symposium for space simulations, March 13–19, RASC, Kyoto University, Uji, Kyoto, Japan, 1997.
16. Klumov B. and B. Lembege, Full particle modeling of plasma cloud expansion into ambiant magnetic field, Advances in Space Res, (Proceedings of 31st COSPAR meeting), in press, 2002.
17. Langdon A.B. and B. F. Lasinski, Methods of Computational Physics, Academic Press, New-York, p. 16, 1976.
18. Lembège B., Numerical simulations of collisionless shocks, in Physical Processes in Hot Cosmic Plasmas, Eds. W. Brinkman, A.C. Fabian and F. Giovannelli, NATO ASI Series, Kluwer Academic Publishers, pp. 81–139, 1990.
19. Lembège B. and J. M. Dawson, Self-consistent study of a perpendicular collisonless and nonresistive shock, Phys. Fluids, 30,1767–1788, 1987a.
20. Lembège B. and J. M. Dawson, Plasma heating through a supercritical oblique collisionless shock, Phys. Fluids, 30, 1110-1114, 1987b.
21. Lembège B. and P. Savoini, Nonstationarity of a two-dimensional quasiperpendicular supercritical collisionless shock by self-reformation, Phys. Fluids, 4, 11, 3533–3548, 1992.
22. Lembège B., S.N. Walker, P. Savoini, M.A. Balikhin, V. Krasnosel'skikh, The spatial size of electric and magnetic field gradient in a simulated shock, Adv. Space Res., (Proceedings of 1998 COSPAR meeting, Nagoya), 24, 109–112, 1999.
23. Lembège B. et Simonet, Hybrid and particle simulations of an interface expansion and of collisionless shock: a comparative and quantitative study, Phys. Plasmas, 8,9, 3967–3981, 2001.
24. Lembège B. and P. Savoini, Formation of reflected electrons bursts by the nonstationarity and nonuniformity of a collisionless shock front, J. Geophys. Res., 107, A3, 4-1 to 4-18, 2002.
25. Lembège B., P. Savoini, M. Balikhin, S. Walker, and V. Krasnosselskikh, Nonadiabatic electron dynamics through a quasi-perpendicular collisionless shock, J. Geophys. Res., submitted to, 2002.
26. Leroy M. M., C. C. Goodrich, D. Winske, C. S. Wu, and K. Papadopoulos, Simulations of perpendicular bow shocks, Geophys. Res. Lett., 8, 1269–1272, 1981.
27. Leroy M. M., D. Winske, C. C. Goodrich, C. S. Wu, and K. Papadopoulos, The structure of perpendicular bow shocks, J. Geophys. Res., 87, 5081–5094, 1982.
28. Liewer P.C., V.K. Decyk, J. M. Dawson, and B. Lembege, Numerical studies of electron dynamics in oblique quasi-perpendicular collisionless shock waves, J. Geophys. Res., 96, 9455, 1991.

29. Ohsawa Y, Plasma heating by a magnetosonic wave through resonant ion acceleration, J. Phys. Soc. Japan, 57, 929–940, 1985.
30. Ohsawa Y, Resonant ion acceleration by collisionless magnetosonic shock wave propagating obliquely to a magnetic field, J. Phys. Soc. Japan, 54, 1657–1660, 1985.
31. Ohsawa Y., Strong ion acceleration by a collisionless magnetosonic shock wave propagating perpendicularly to a magnetic field, Phys. Fluids, 328, 2130–2136, 1985.
32. Omidi N., and D. Winske, Kinetic structure of slow shocks: effects of the electromagnetic ion/ion cyclotron instability, J. Geophys. Res., 97, A10, 14801–14821, 1992.
33. Papadopoulos K., Microinstabilities and anomalous transport, in Collisionless shocks in the heliosphere: a tutorial review, Edt. G. Stone and B.T. Tsurutani, Geophysical Monograph 34, American Geophysical Union, Washington, D.C., 1985.
34. Savoini P. and B. Lembège, Electron dynamics in two-and one-dimensional oblique supercritical collisionless magnetosonic shocks, J. Geophys. Res., 99, A4, 6609-6635, 1994.
35. Savoini P. and B. Lembège, Heating and acceleration of electrons through the whistler precursor in 1D and 2D oblique shocks, Adv. Space Rev., 15, No 8/9, 235-238,(Proceedings of the D2.1 Symposium of COSPAR Scientific commission D, 30th COSPAR meeting, Germany), 1995.
36. Savoini P. and B. Lembège, Full curvature effects of a collisionless shock, Adv. Space Res., (Proceedings of 1998 COSPAR meeting, Nagoya), 24,13–22, 1999.
37. Savoini Ph. and B. Lembège, Two-dimensional simulations of a curved shock: self-consistent formation of the electron foreshock, J. Geophys. Res., 106, A7, 12975–12992, 2001.
38. Shimada N. and M. Hoshino, Strong electron acceleration at high Mach number shock waves: simulation study of electron dynamics, Astrophy.J., 543, L76-L71, 2000.
39. Tidman and Krall, Shock waves in collisionless plasmas, Wiley-Interscience Series in Plasma Physics, 1971.
40. Thomas V. A. and D. Winske, Two dimensional hybrid simulation of a curved bow shock, Geophys. Res. Lett., 17, 1247–1250, 1990.

Simulation of Electron Beam Instabilities and Nonlinear Potential Structures

Yoshiharu Omura[1], Takayuki Umeda[1], and Hiroshi Matsumoto[1]

Radio Science Center for Space and Atmosphere, Kyoto University, Uji, Kyoto 611-001, Japan

Abstract. This paper gives a tutorial for the electrostatic instabilities induced by electron and/or ion beams in space plasmas. After a brief review of the linear theory of the electron/ion beam instabilities, we describe various types of nonlinear processes reproduced by electrostatic particle simulations. We first study the basic nonlinear processes in a uniform system with periodic boundaries, and then we study the beam plasma interactions in a nonuniform system with open boundaries. Importance of ion dynamics in the nonlinear evolution of the electron beam instabilities are discussed based on the simulations with different distributions of ions.

1 Introduction

Electron beams are formed in various processes in space plasmas, such as particle reflection at shocks, magnetic reconnection, inductive electric field, and parallel electric field of kinetic Alfven waves. Electron beams cause strong electrostatic instabilities, which lead to excitation of Langmuir waves, electrostatic solitary waves (ESW) or ion acoustic waves through various nonlinear processes. Electromagnetic waves such as whistler mode waves are also excited by electron beams directly via wave-particle interactions or indirectly through wave-wave interactions. In this tutorial, however, we focus our attention on electrostatic potential structures along the magnetic field, which are often observed in the plasma sheet boundary layer, the auroral region and the magnetosheath. In these regions, electron beams are also observed as enhanced non-thermal fluxes with flat-top diffused distribution functions.

The purpose of this article is to provide an introduction to nonlinear processes associated with these observations of electrostatic waves and particles in space. In the following sections, we first study linear dispersion properties of electron beam instabilities, and then dynamics of resonant particles and formation of ESW as reproduced by a simple electrostatic particle simulation. We also study decay of ESW in a uniform periodic system, spatial evolution of a localized electron beam in an open system, and coupling with oblique low frequency modes in a two-dimensional system. These nonlinear electrostatic wave phenomena are readily reproduced by particle-in-cell simulations where electrostatic field are solved either by solving Poisson's equation or by solving a full set of Maxwell's equations. Detailed descriptions of the numerical techniques are found in chapters for KEMPO1 and TRISTAN codes in a simulation book edited by Matsumoto and Omura [9]. The book and simulation codes can be downloded

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 79–92, 2003.

from the website: http://www.terrapub.co.jp/. The following sections describe essential physical processes found in simulation runs where electron/ion beams are assumed as the initial condition.

2 Linear Dispersion Relation

We assume a few species of electrons and ions drifting along the static magnetic field. Particle distribution functions are defined in the three-dimensional velocity space. For simplicity, however, we neglect electric and magnetic field perturbations in the directions perpendicular to the static magnetic field. By integrating the velocity distribution function with two velocity components perpendicular to the magnetic field, we obtain a reduced velocity distribution function $f_s(v, x, t)$ of the parallel velocity component v. The kinetic description of electron beam instability is given by the dispersion relations derived from the Vlasov equation for the species "s",

$$\frac{\partial f_s}{\partial t} + v\frac{\partial f_s}{\partial x} - \frac{q_s}{m_s}\frac{\partial \phi}{\partial x}\frac{\partial f_s}{\partial v} = 0 \tag{1}$$

and Poisson's equation,

$$\frac{\partial^2 \phi}{\partial x^2} = -\frac{1}{\epsilon_o}\sum_s q_s \int_{-\infty}^{\infty} f_s \mathrm{d}v \tag{2}$$

where ϕ is the electrostatic potential and the x-axis is taken along the magnetic field. It should be noted that there arises no effect of the magnetic field on electrostatic interactions in the one-dimensional system taken along the magnetic field.

We linearize the Vlasov equation by separating the distribution function into unperturbed and perturbed components

$$f_s(v, x, t) = f_{so}(v) + f_{s1}(v, x, t) \tag{3}$$

We assume a Maxwellian velocity distribution function for the unperturbed component.

$$f_{so}(v) = \frac{n_{so}}{\sqrt{2\pi}V_{ts}}\exp(-\frac{(v - V_{ds})^2}{2V_{ts}^2}) \tag{4}$$

where n_{so} is an unperturbed number density, and V_{ts} and V_{ds} are thermal and drift velocities, respectively. Applying Fourier and Laplace transforms to the Vlasov and Poisson equations in space and time, respectively [e.g., Ref. [11]], we obtain the following dispersion relation of electrostatic waves with wave vectors parallel to the static magnetic field.

$$D(k, \omega) \equiv 1 - \sum_s \frac{\Pi_s^2}{k^2}\int_L \frac{\mathrm{d}g_{so}/\mathrm{d}v}{v - \omega/k}\mathrm{d}v = 0 \tag{5}$$

where Π_s is the plasma frequency of particle species "s" given by

$$\Pi_s = \sqrt{\frac{n_{so}q_s^2}{\epsilon_o m_s}}, \tag{6}$$

and $g_{so}(v)$ is a normalized unperturbed velocity distribution function given by $f_{so}(v)/n_{so}$. The wavenumber is assumed to be a positive real variable and the frequency is a complex variable, i.e., $\omega = \omega_r + i\gamma$. The integration over the velocity is taken along the Landau contour [7]. When the thermal velocity is comparable to the phase velocity, the dispersion relation yield a finite imaginary part, i.e.,

$$D_r(k,\omega) + i\ D_i(k,\omega) = 0 \tag{7}$$

If $|\gamma| << |\omega_r|$, we can obtain an approximate growth rate,

$$\gamma = -\frac{D_i(k,\omega_r)}{\partial D_r(k,\omega_r)/\partial\omega_r} \tag{8}$$

An instability ($\gamma > 0$) due to the imaginary part of D_i is called a "resistive instability" .

When the thermal velocities are small enough, i.e.,

$$\frac{\omega}{k} - V_{ds} >> kV_{ts}, \tag{9}$$

the dispersion relation is simplified to the following form

$$1 = \sum_s \frac{\Pi_s^2}{(\omega - kV_{ds})^2} \tag{10}$$

It is noted that the dispersion relation has no imaginary part. An instability due to a positive γ of a solution of the above equation is called a "reactive instability" . A typical example of reactive instabilities is a bi-stream instability with two cold electron beams of approximately equal densities as shown in the left panels of Figs. 1(a) and 1(b). Let us assume the densities of the electron beams are equal, and the relative drift velocity between them is V_d. Inserting $\Pi_1 = \Pi_2 = \Pi_e/\sqrt{2}$ into (10), and neglecting the term of the background ions, we find that the maximum growth rate $\gamma_{max}$is $2^{-3/2}\Pi_e$ at $k = 3^{1/2}2^{-3/2}\Pi_e/V_d$.

Another example is the Buneman instability [3] caused by a single electron population with a finite drift velocity with respect to the background ions. In Fig. 2, we plotted linear growth rates of the Buneman instability for different thermal velocities of electrons. In the case of reactive instabilities with cold electrons and ions, the dispersion relation takes the following form rewritten from (10).

$$1 = \frac{\Pi_e^2}{\omega^2} + \frac{\Pi_i^2}{(\omega - kV_d)^2} \tag{11}$$

where the cold electrons are at rest and ions are drifting with V_d. Since $\Pi_i^2 << \Pi_e^2$, the dispersion relation is also for the electron weak-beam instability illustrated in Fig. 1(c), if we assume Π_i represents the plasma frequency of the weak electron beam. Assuming $\alpha \equiv \Pi_i^2/\Pi_e^2 << 1$, we can obtain the maximum growth rate $\gamma_{max} = 3^{1/2}2^{-4/3}\alpha^{1/3}\Pi_e$ with $k = \Pi_e/V_d$ (see e.g. Ref. [14]). Noting that $\alpha = m_e/m_i$ for the Buneman instability, we find $\gamma_{max} = 0.056, 0.15, 0.24$ for

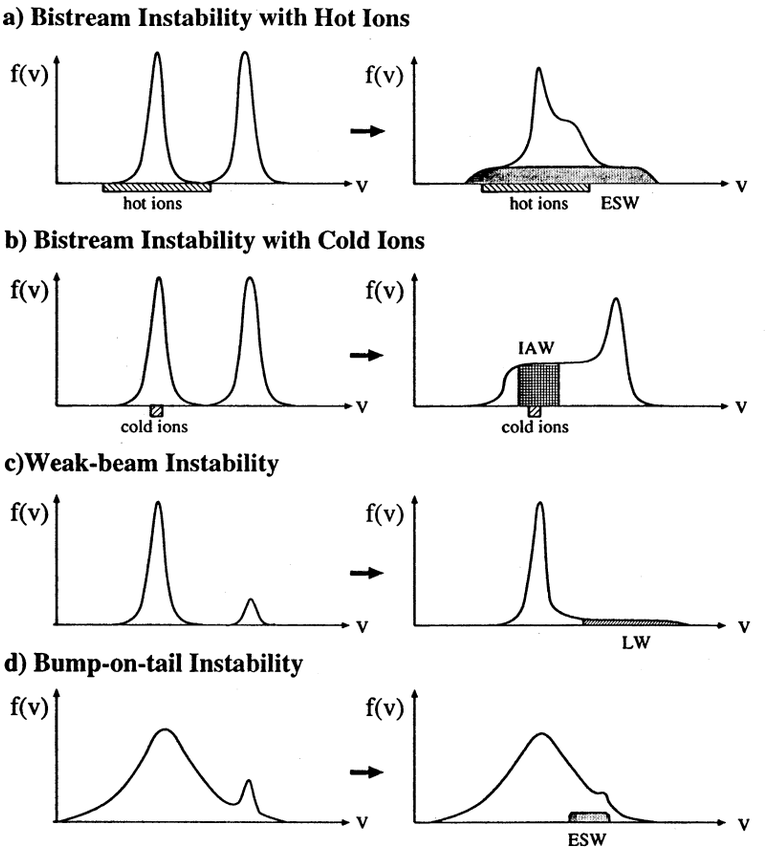


Fig. 1. Initial and final velocity distribution functions of (a) bi-stream instability with hot ions, (b) bi-stream instability with cold ions, (c) weak-beam instability, (d) bump-on-tail instability. [after Ref. [12]]

mass ratios $(m_i/m_e) = 1836, 100, 25$, respectively. The reduced mass ratios are often used in full particle simulations to reproduce coupling processes of electron and ion dynamics with a limited computer capability.

In the Buneman instability, a wave mode with the maximum growth rate grows to a large level to form large potentials that can trap the whole population of electrons. Electrons are diffused over a wide range of velocity around the drifting ions, whereas an electron beam in the weak-beam instability is diffused over a relatively small range of velocity around a velocity slightly smaller than its initial velocity. As the thermal velocity of electrons increases, the growth rate becomes small as shown in Fig. 2. The nature of the instability becomes resistive rather than reactive.

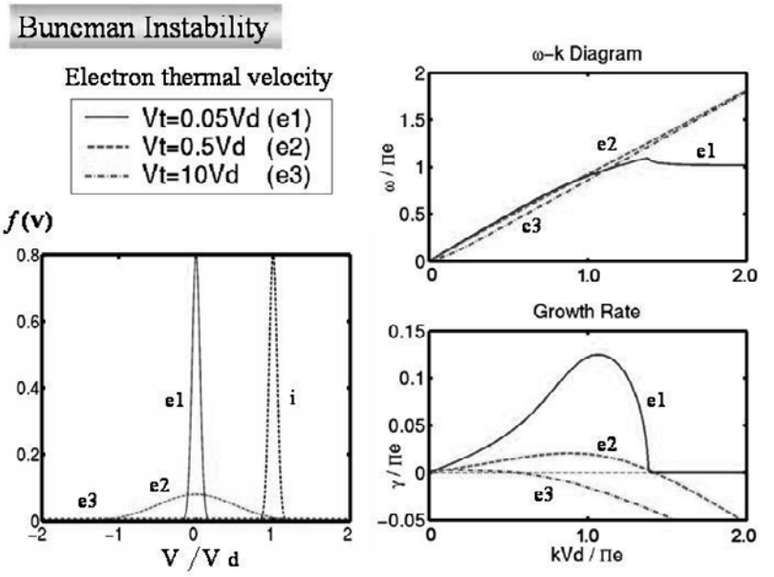


Fig. 2. Linear dispersion relations of Buneman Instability for different thermal velocity of electrons. A reduced mass ratio $m_i/m_e = 100$ is assumed.

The maximum growth rates of resistive instabilities depend on the density of the electron beams and the thermal velocity of the background electrons and/or ions. More specifically, the gradient of the velocity distribution function at the phase velocity of a growing wave determines the growth rate. A typical example is the bump-on-tail instability as illustrated in Fig. 1(d). A wave mode with the maximum growth rate grows from a thermal noise level to a level that traps the resonant electrons diffusing the electron beam effectively. This is a coherent process due to the dynamics of phase bunched electrons as described in the following sections.

When the growth rate is very small ($\gamma << \omega$ and $\gamma << kV_{ts}$), the quasi-linear diffusion [11] takes place to make the velocity distribution function marginally stable. The diffusion is an incoherent process induced by independent waves with random phases. In most of beam instabilities starting from thermal fluctuations, the quasi-linear diffusion is only found after the saturation and decay of the most unstable mode. In particle simulations, thermal fluctuations are especially enhanced because of the limited number of superparticles in the Debye length. The enhanced fluctuations can also cause a diffusion similar to the quasi-linear diffusion.

3 Nonlinear Trapping of Particles

When an electron beam instability saturates after a sufficient growth time ($\gamma_{max}t > 1$), most of the electrons forming the beam are trapped by coherent electrostatic potentials formed by the dominant wave mode with the maximum linear growth rate. The trajectories of the trapped electrons are described by the equations of motion under a wave with wavenumber k, frequency ω and wave amplitude E_w.

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{q_s}{m_s} E_w \sin\left(kx - \omega t + \zeta_o\right) \tag{12}$$

$$\frac{\mathrm{d}x}{\mathrm{d}t} = v \tag{13}$$

Taking a frame of reference moving with the wave phase velocity as a variable of the velocity,

$$\theta = k(v - \frac{\omega}{k}) = kv - \omega \tag{14}$$

and defining a phase $\zeta = kx - \omega t + \zeta_o + \pi$ for a positive charge ($q_s > 0$) and $\zeta = kx - \omega t + \zeta_o$ for a negative charge ($q_s < 0$), we obtain the equation of a pendulum.

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = -\omega_t^2 \sin\zeta \tag{15}$$

$$\frac{\mathrm{d}\zeta}{\mathrm{d}t} = \theta \tag{16}$$

where ω_t is the trapping frequency given by

$$\omega_t = \sqrt{\frac{k|q_s|E_w}{m_s}} \tag{17}$$

Integrating (15) and (16), we obtain an equation of a particle trajectory in (θ, ζ).

$$\theta^2 = 2\omega_t^2 \cos\zeta + C \tag{18}$$

where C is a constant corresponding to a specific trajectory.

Resonant particles with velocities close to the wave phase velocity oscillate around a stable equilibrium point at $(\theta, \zeta) = (0, 0)$ with the trapping frequency. The saddle point of resonant particle trajectories is located at $(\theta, \zeta) = (0, \pm\pi)$, which gives a separatrix of the trapping region. The maximum values of θ of the trapping region is given by $kV_t = 2\omega_t$, where V_t is called a trapping velocity. In the presence of a large electrostatic wave, the velocity distribution function of the electron beam becomes flat over the range of $V_r - V_t < v < V_r + V_t$. Resonant electrons are diffused over the velocity range, but it has a distinct structure in the velocity phase space. The trapped particles oscillate fast in the space around $(\theta, \zeta) = (0, 0)$ with frequencies close to the trapping frequency ω_t, while particles close to the separatrix of the trapping region move slowly especially near the saddle point (Fig. 3). The modulation in the velocity results in

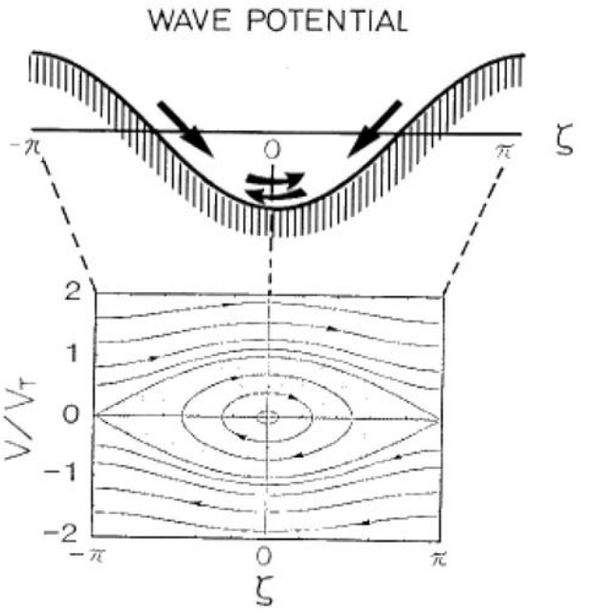


Fig. 3. Schematic illustration of particle trapping in a wave potential well [8].

the density modulation. Especially when the resonant electrons become trapped under a growing wave, most of the particles remain just inside the boundary of the separatrix making the center of trapping oscillation $(\theta, \zeta) = (0, 0)$ a void region, which we call an "electron hole" for electrons and an "ion hole" for ions. An electron hole corresponds to a positive electrostatic potential, while an ion hole is a negative potential.

4 Coalescence of Electron Holes

Most of the electron holes, whose drift velocities are approximately the same, coalesce with each other by converting a part of the electrostatic energy to the thermal energy of trapped and untrapped resonant electrons. In Fig. 4, we plotted a series of phase plots of electron holes found in a long time evolution of bi-stream instability with warm ions as illustrated in Fig. 1(a). We assume a one-dimensional system consisting of 1024 grid points and periodic boundaries. We put two electron beams with equal densities and a neutralizing thermal ions at rest with one of the electrons. The other electrons are drifting with a velocity V_d. The thermal velocities of both electron beams are $V_{te} = 0.05V_d$, while that of ions is $V_{ti} = 0.1V_d$. We assumed a reduced mass ratio $m_i = m_e = 100$. In order to suppress the thermal fluctuation, we used 10,240 superparticles per cell for each species.

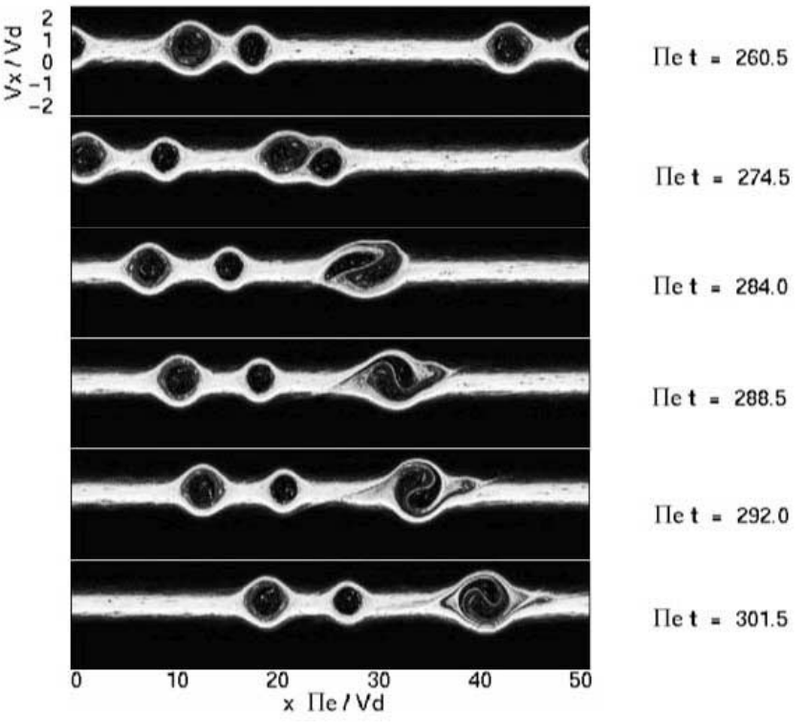


Fig. 4. Phase plots of two electron holes coalescing with each other to form a larger electron hole. Some of the trapped particles are detrapped from the trapping region, releasing energy from the coalesced electron holes to establish a new state of the BGK equilibrium.

When the excited electrostatic potentials are on a normal mode without dispersion, the potential structures travel with the same phase velocity. The adjacent electron holes attract each other [4] and they coalesce with each other to form larger electron holes. Through coalescence there arise isolated potentials, which are close to the BGK equilibrium, which is a time-independent solution of Vlasov-Poisson equations [1].

By taking a frame of reference moving with one of the electron holes, where we can assume that $\partial/\partial t = 0$, we obtain from (1),

$$v\frac{\partial f_s}{\partial x} - \frac{q_s}{m_s}\frac{\partial \phi}{\partial x}\frac{\partial f_s}{\partial v} = 0 \tag{19}$$

We define an energy variable

$$W_s \equiv \frac{1}{2}m_s v^2 + q_s\phi(x) \tag{20}$$

From (19) we find that an arbitrary function of the energy variable $f_s(x, v) = F(W_s)$ can be at the BGK equilibrium satisfying (19) and (2). Denoting $F(W_s)$

as F_s^+ and F_s^- respectively for positive ($v > 0$) and negative ($v < 0$) velocities, we have the particle density expressed as

$$n_s(x) = \int_{-\infty}^{0} f_s(x,v)\mathrm{d}v + \int_{0}^{\infty} f_s(x,v)\mathrm{d}v \tag{21}$$

$$= \int_{q_s\phi}^{\infty} \frac{F_s^- + F_s^+}{\sqrt{2m_s(W_s - q_s\phi)}}\mathrm{d}W_s \tag{22}$$

The kinetic energy density is also expressed as

$$U_s(\phi) = \int_{-\infty}^{\infty} \frac{1}{2} m_s v^2 f_s \mathrm{d}v$$

$$= \frac{1}{2m_s} \int_{q_s\phi}^{\infty} (F_s^- + F_s^+)\sqrt{2m_s(W_s - q_s\phi)}\mathrm{d}W_s \tag{23}$$

Combining these expressions, we obtain a useful relation valid for the BGK equilibrium [5].

$$\frac{\mathrm{d}U_s}{\mathrm{d}\phi} = -\frac{q_s}{2} n_s \tag{24}$$

Poisson's equation (2) is rewritten as

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} = \frac{2}{\epsilon_o}\frac{\mathrm{d}}{\mathrm{d}\phi}\sum_s U_s \tag{25}$$

We multiply the above equation by $\mathrm{d}\phi/\mathrm{d}x$ and integrate over x to obtain

$$\frac{\epsilon_o E_x^2}{2} = 2\sum_s (U_s - U_{so}) \tag{26}$$

The integration constant U_{so} corresponds to an unperturbed kinetic energy density at infinity where $E_x = 0$. The equation derived above reveals that the electrostatic energy density is exactly twice as large as the perturbation of the kinetic energy at the BGK equilibrium [5].

At the moment of coalescence of two electron holes as shown in Fig. 4, there arises a stronger electrostatic field because of overlapping of the potentials, accelerating trapped electrons in the potentials. Some of the trapped electrons become untrapped and move away from the coalesced potential. Thus both electrostatic and kinetic energies of the two electron holes are lost to form a new electron hole at the BGK equilibrium satisfying (26). Thus, the coalescence is an irreversible process.

5 Decay of Electron Holes

In a system where normal modes with different phase velocities can exists stably without much damping, nonlinear potentials excited by the electron beam

instabilities decay into these normal modes with different phase velocities. An example is a weak beam instability where an electron beam with a few percent of the total density is injected into the cold background component as shown in Fig. 1(c). Because of the absence of electrons between the beam drift velocity and that of the thermal plasma, Langmuir waves with phase velocities less than the beam drift velocity are subsequently excited through the quasi-linear diffusion process. The coherency of the potentials decreases, and the phases of the normal modes become random. There occurs the quasi-linear diffusion which leads to a formation of plateau in the velocity distribution function of electrons.

Another example where electron holes decay into other waves is a bi-stream electron instability with cold ions as shown in Fig. 1(b). The bi-stream electron beams cause a very strong reactive instability, generating large electron holes which make the eventual thermal velocity of the electrons very large. By decreasing the ion thermal velocity V_{ti} from $0.1V_d$ to $0.05V_d$, we obtained very different results of long-time evolution of the bi-stream instabilities. In the presence of the cold ions, ion acoustic waves exist, being free from the ion Landau damping. The large electron holes decay into the ion acoustic waves . If the ions are warm, the electron holes exist stably because the ion acoustic waves are subject to the ion Landau damping. Thus the parameters of ions become critically important in the long time evolution of the electron holes.

The importance of the ion dynamics can be found in a bi-stream electron instability with an warm ion beam placed between the two electron beams in the velocity space. After the initial formation of large electron holes, we find the electron holes are split into two smaller electron holes moving forward and backward with respect to the ion beam at rest. As seen in this example, the electron holes moving with sufficiently large drift velocities with respect to each other do not coalesce through a collision. In each frame of the reference the drift energy of the other beam is too large to be trapped by the potential well.

6 Electron Beam Instability in Open Systems

In a system with periodic boundary, beam instabilities take place uniformly in space, and evolution of the instability can be traced infinitely in time. In real space plasmas, sources of electron beams or acceleration regions are localized in space. We study beam instabilities started from a localized point source in this section. We set up a one-dimensional simulation region consisting of 10,240 grid points with open boundaries where outgoing waves and particles are absorbed as if there is no boundary. Thermal particles are injected from both boundaries to maintain the thermal populations of ions and electrons at rest. From the left boundary we inject a flux of electrons with a drift velocity V_d and a thermal velocity $V_b = 0.05$ so that 5 % of the total electrons from an electron beam in the velocity phase space. The thermal velocity of the background thermal electrons is $V_{te} = 0.5V_d$. We also put a neutralizing cold ions with a thermal velocity $V_{ti} = 0.005V_d$ and with a reduced mass ratio $m_i/m_e = 100$. The number of superparticles is 1600 per cell for each of electrons and ions.

The initial velocity and position of a particle injected from the left boundary with a positve velocity are given by using two uniformly distributed random numbers $0 < R_1, R_2 < 1$ as shown in the following procedure [2]. The particles are injected with a flux velocity distribution function given by $vf(v)$, where $f(v)$ is a velocity distributuion function for a species of particles.

$$v = F^{-1}(R_1) \tag{27}$$

$$x = R_2 v \frac{\Delta t}{2} \tag{28}$$

where

$$R_1 = F(v) \equiv \frac{\int_0^v v' f(v') \mathrm{d}v'}{\int_0^{v_{max}} v' f(v') \mathrm{d}v'}. \tag{29}$$

When a computer simulation is started, the electron beam is continuously injected from the left boundary of the simulation box into the background homogeneous plasma. We inject particles with the constant flux, while exiting particles are removed at the open boundaries. The electrostatic fields at the open boundaries are determined by the distribution of charged particles without any special numerical treatments. We first compute the charge density distribution at the open boundaries. Then we solve Poisson's equation to obtain the electrostatic potential and the electrostatic field.

As we found in the uniform periodic system an electron beam on a high energy tail of the warm plasma can generate electron holes as shown in Fig. 5. The electron holes impinge into the unperturbed plasma without significant disturbances and propagate stably for a long distance. This implies that the electron holes are formed essentially by dynamics of trapped electrons rather than that of untrapped resonant electrons that pass through the electron holes. Contrary to the case of the bi-stream instability with cold ions, the excited electron holes does not decay into ion acoustic waves. This is because the amplitude of the electron holes are small enough to couple with the cold ions at rest.

When the same electron beam is injected into the cold plasma, however, the large electrostatic potentials excited by the weak beam instability rapidly decay into Langmuir waves with smaller phase velocities. Contrary to the uniform system, the top portion of electron beam with the largest positive velocity moves into the unperturbed plasma. The diffused parts with smaller phase velocities follow the top portion with some time delays. This keeps the top part of the electron beam unstable to the weak-beam instability, although the growth rate decreases because of the decreasing density of the beam.

7 Nonlinear Evolution in Two-Dimensional System

In the preceding sections, we have studied beam instabilities in the one-dimensional system taken along the static magnetic field. Since the growth rates of the electron beam instabilities maximize in the parallel direction, the one-dimensional model is valid at the initial stage of the nonlinear wave-particle interaction. In

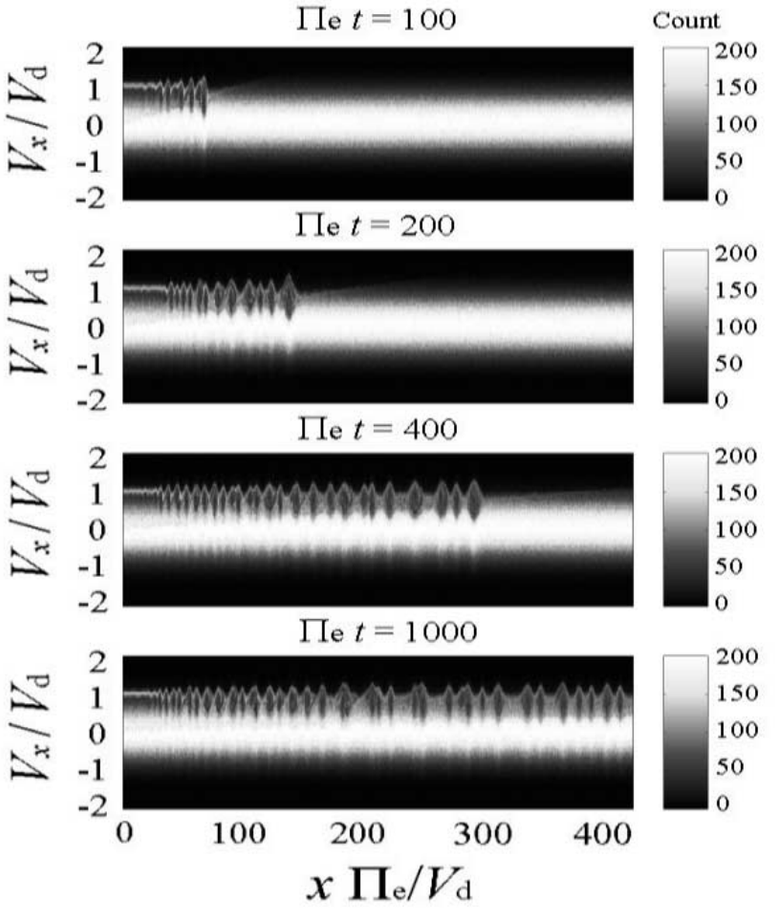


Fig. 5. Spatial evolution of the bump-on-tail instability in the one-dimensional open system [13]. The $x - v_x$ phase diagrams at $\Pi_e t = 100$, 200, 400 and 1000, respectively.

a two-dimensional periodic system with a spatial dimension perpendicular to the static magnetic field, the one-dimensional ESW can couple with the wave modes propagating in oblique directions when the following resonance condition is satisfied.

$$\frac{\omega}{k\cos\theta} = V_d \tag{30}$$

where θ is an angle between the wavenumber vector $\boldsymbol{k}$ and the static magnetic field $\boldsymbol{B}$. The velocity V_d is the drift velocity of ESW. The left-hand-side is the phase velocity of the oblique mode parallel to the static magnetic field. (See Fig. 6).

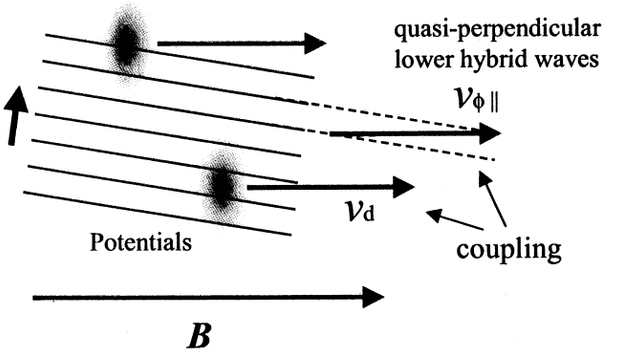


Fig. 6. Schematic drawing of coupling between two-dimensional potentials drifting parallel to the magnetic field **B** and low frequency waves propagating in a oblique direction [10].

When the resonance condition is satisfied, there develops a modulation of the one-dimensional potential structures excited by the electron beam instability, leading to a formation of two-dimensional potential structures [10]. The solitary potentials also work as fast moving charged particles, from which obliquely propagating whistler mode waves are generated [6]. In an open system where the source of the electron beam is localized, however, the top portion of the electron beam impinging into an unperturbed plasma is free from the oblique modes, because their group velocities are much smaller than that of the electron beam. Only in the localized region close to the source of the electron beam, we find coupling of ESW with the lower hybrid waves [15]. The localized two-dimensional and three-dimensional structures of solitary waves and their efficiency for electromagnetic wave radiation are subjects of the on-going studies.

Acknowledgements

We thank H. Kojima, H. Usui, V. Krasovsky, and K. Ninomiya for their discussions on the simulation studies of the electrostatic solitary waves. The computer simulations were done on the AKDK computer system at Radio Science Center for Space and Atmosphere, Kyoto University. The present study is partially supported by Grant-in-Aid 13440143 of Japan Society for the Promotion of Science.

References

1. Bernstein, I. B., J. M. Greene, and M. D. Krushkal, Exact nonlinear plasma oscillations, Phys. Rev., 108, 546, 1957.
2. Birsall, C. K., and A. B. Langdon, Plasma Physics via Computer Simulation, McGraw-Hill, 1985.
3. Buneman, O., Dissipation of currents in ionized media, Phys. Rev., 115, 503, 1959.
4. Krasovsky, V. L., H. Matsumoto, and Y. Omura, Interaction of small phase density holes, Physica Scripta, 60, 438, 1999.

5. Krasovsky, V. L., H. Matsumoto, and Y. Omura, Electrostatic solitary waves as collective charges in a magnetospheric plasma, J. Geophys. Res., in press, 2002.
6. Goldman, M. V., M. M. Oppenheim, and D. L.Newmann, Nonlinear two-stream instabilities as an explanation for auroral bipolar wave structures, Geophys. Res. Lett., 26, 1821, 1999.
7. Landau, L. D., On the vibrations of the electronic plasma, J. Phys. (USSR), 10, 25, 1946.
8. Matsumoto, H., Coherent nonlinear effects on electromagnetic wave-particle interactions, Space Sci. Rev., 42, 429, 1985.
9. Matsumoto, H., Y. Omura, Computer Space Plasma Physics, Terra Scientific Pub. Co., Tokyo, 1993.
10. Miyake, T., Y. Omura, and H. Matsumoto, Electrostatic particle simulations of solitary waves in the auroral region, J. Geophys. Res., 105, 23239, 2000.
11. Nicholson, D. R., Introduction to plasma theory, John Wiley & Sons, 1983.
12. Omura, Y., H. Matsumoto, T. Miyake and H. Kojima, Electron Beam Instabilities as Generation Mechanism of Electrostatic Solitary Waves in the Magnetotail, J. Geophys. Res., 101, 2685, 1996.
13. Omura, Y., H. Kojima, T. Umeda, and H. Matsumoto, Observational evidence of dissipative small scale processes: Geotail spacecraft observation and simulation of electrostatic solitary waves, Astrophys. Space Sci. J., 277, 45, 2001.
14. O'Neil, T. M., J. H. Winfrey, and J. H. Malmberg, Nonlinear Interaction of a Small Cold Beam and a Plasma, Phys. Fluids, 14, 1204, 1971
15. Umeda, T., Y. Omura, H. Matsumoto, and H. Usui, Formation of electrostatic solitary waves in space plasmas: Particle simulations with open boundary conditions, J. Geophys. Res., in press, 2002.

Kinetic Simulation of Inhomogeneous Plasma with a Variable Sized Grid System

David Schriver

University California Los Angeles, Los Angeles CA 90095-1567, USA

Abstract. Space plasmas are usually inhomogeneous and irregular. Particle in cell (PIC) plasma simulations, however, often consider the plasma to be homogeneous and periodic. Here non-periodic PIC simulations with inhomogeneities in the density, temperature and background magnetic field are considered. Boundary conditions and a variable sized grid system are discussed, along with an application to a magnetospheric plasma physics problem in the auroral zone.

1 Introduction

Particle in cell (PIC) simulations have been used to simulate plasmas since the early 1960's [10]. The basic idea behind PIC simulations is to follow charged particles self-consistently in some region of space as they move according to the Lorentz force equation and the fields evolve per Maxwell's equations. The practice and use of standard PIC simulations has been well documented over the years [7]. In most cases, to resolve the plasma a finite regular grid and time step are used.

PIC codes are not the only way to simulate space plasmas and there are numerous other types of self-consistent simulations that treat the plasma according to some approximation. This includes fluid codes using the magnetohydrodynamic (MHD) method [22], hybrid models whereby ions are treated as particles and electrons as a fluid [42], and Vlasov codes whereby the distribution functions evolve according to the Vlasov equation [20],[32]. A discussion of some of these codes and their applications to space plasmas can be found in the proceedings of the Second and Third International School of Space Simulations [44],[45].

Here only PIC simulations will be considered, and in particular the workings and use of inhomogeneous and non-periodic codes will be discussed. Although some basic tenets of PIC simulations will be presented here, the focus will be on the application of variable grid spacing and non-periodic boundary conditions. For a complete description of basic PIC codes, the reader is referred to [7],[11] and the references cited there.

This paper is structured as follows. In the next section is a discussion of why inhomogeneous and non-periodic systems are needed. In Sect. 3 a scheme for incorporating variable grids into PIC simulations is presented, followed in Sect. 4 by a discussion of non-periodic boundary conditions. An example of a PIC simulation code with variable grids and non-periodic boundary conditions applied to the auroral zone is given in Sect. 5. Conclusions and future directions are found in Sect. 6.

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 93–108, 2003.

2 Why Non-periodic and Inhomogeneous Systems?

Particle in cell simulations are an excellent tool for examining instabilities and wave-particle interactions in a plasma. Since the Lorentz equation is used to push many (up to several million) charged particles according to either the full or a reduced set of Maxwell's equations, in principle all of the normal plasma wave modes, as well as linear and non-linear interactions are naturally included. Many applications consider a beam driven instability or some other type of anisotropy in a uniform, periodic system. Finite grid systems and time steps are used in most PIC codes and a periodic system allows the equations to be solved in Fourier transform space, which in general is a fast, relatively easy way of simulating a plasma. Such simulations can allow an identification of plasma wave modes excited in the system and an understanding of wave saturation mechanisms.

Satellite observations in space, however, usually find that the plasma is not uniform. There are often gradients and shear flows such as that found at the magnetopause boundary layer between the solar wind and the Earth's magnetosphere [1],[12],[27] at the plasma sheet boundary layer in the magnetotail [13],[15],[16] and at times in the plasma sheet itself [4],[6],[24], to name only a few. These gradients can occur in both particle and field profiles. Another, somewhat more gradual non-uniform system occurs where the cool, dense, highly magnetized plasma of the ionosphere near the Earth gives way to the warm, tenuous, weakly magnetized plasma in the magnetotail plasma sheet at high altitudes.

Since PIC simulation scale lengths are in most cases determined by the electron Debye length (e.g., [7]), system sizes are often relatively small (i.e., kilometers) compared to observed spatial scales ($\sim$ one Earth radii, where 1 $R_E = 6371$ km). In many cases the use of periodic, uniform systems that simulate a very small region of space is a reasonable approximation. But the rapid increase in computing speed and relatively large amounts of memory now available have allowed simulations with more grid points and much larger system sizes to be carried out. Since satellite observations show the magnetosphere to be spatially and temporally non-uniform over R_E spatial scales (or less) and since the various regions depend on one another dynamically, large-scale PIC simulations must be non-uniform and non-periodic.

3 Simulations with Irregular Grid Systems

In most Cartesian PIC simulation systems the grid spacing is regular, i.e., evenly spaced. In non-Cartesian systems the grid spacing can be irregular (e.g., [7]). Here we consider a system that has a gradual variation in density, temperature, and magnetic field over thousands of kilometers such as in the auroral zone. To examine the plasma physics of such a system, an irregular grid, non-periodic simulation is employed. Observed plasma conditions in the auroral region determine the simulation system. The auroral zone is characterized by a transition from the relatively low density, warm, magnetotail plasma in a weakening dipole magnetic field at high altitudes to a region near the Earth in the ionosphere

where the plasma is cold, dense and the magnetic field is strong. The variation in plasma density and temperature can be several orders of magnitude from the ionosphere located at a few hundred kilometers altitude to the magnetotail, which begins at an altitude of several Earth radii and beyond. The auroral zone sits between these regions at a few thousand to tens of thousands of kilometers altitude and is distinguished by field-aligned and transverse particle acceleration, intense wave activity of various kinds, and quasi-static electric fields. A source of free energy in the form of field-aligned currents, Alfven wave Poynting flux, and particle beams flow from the magnetotail towards the ionosphere and drive the auroral region's dynamic behavior. To model this region as a whole, a PIC simulation must stretch from the high altitude magnetosphere to the low altitude ionosphere and include a source of free energy that flows into the system.

An example of a PIC simulation model of the auroral zone is described here (results from the code are presented in Sect. 5). To model a large portion of the auroral zone a spatial system with a length as large as or greater than an Earth radii (> 6300 km) is needed. A one-dimensional system will be considered with a system axis aligned along the Earth's dipole magnetic field, which near the Earth is approximately equivalent to altitude. Figure 1 shows schematically the location of the simulation system with respect to the Earth along with the magnetic field profile assumed. To achieve a system length of an R_E or more for this plasma configuration, tens of thousands of grid points are used. For the simulations discussed here, 50,000 grid points is the minimum number used and runs with up to 100,000 grid points have been used.

Standard practice in PIC simulations is to divide the spatial system over which the particles move into a number of equally spaced grid cells. Although the particle position is continuous over the system, particle density and current are calculated only at the grid points, which is also where the electric field values are calculated from Poisson's equations. Field values are then extrapolated from the grid to the particle location to push the particles. It has been shown rigorously that in order to avoid non-physical heating and other spurious effects due to the use of finite sized grid cells, the size of each grid should be set approximately equal to the electron Debye length (e.g., [7]). The electron Debye length λ_e is given by:

$$\lambda_e = \sqrt{\frac{k_B T_e}{4\pi n_e q^2}} \tag{1}$$

The parameters are the electron temperature T_e, the electron density n_e, Boltzmann's constant k_B and the charge q. A problem in simulating the auroral zone deals with treating a plasma that is dense and cold, and thus has a small Debye length, in one part of the system, while in another part of the system the plasma is tenuous and warm with a relatively large Debye length. To deal with this a variable grid system has been employed. In the auroral zone the density and temperature, according to observations, depend on altitude (r) approximately as [19]:

$$n(r) = n_0 e^{-(r-r_0)/h} + 17(r-1)^{-1.5} \tag{2}$$

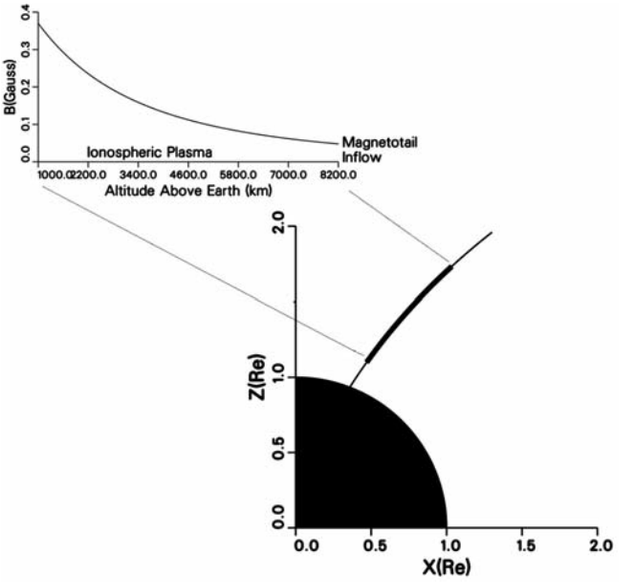


Fig. 1. Schematic representation of a simulation system that models the auroral region. The Earth is shown at the lower right and the thin line represents a 70° magnetic field line. The thicker line shows the region being modeled using the PIC simulation, which is expanded in the upper left part of the figure.

$$T(r) = T_1 e^{r/h_0} + T_0 \tag{3}$$

The different parameters such as r_0, h_0, T_1, T_0, and n_0 are various scale heights, base temperatures and base densities whose exact numerical values are not important for this discussion. Number values based on observations can be found in [19]. The main point to be emphasized here is that the density tends to decrease exponentially with altitude while the temperature increases exponentially with altitude. These combined effects thus lead to an overall increase in the electron Debye length with altitude.

The auroral simulation system is set up such that the size of the first grid cell (denoted by Δ_1) has a length determined by the Debye length at the lowest altitude portion of the system where $\lambda_e(r = r_0) = \lambda_{e0}$. For the simulation runs discussed here the base altitude r_0, is set to 1000 km. The size of the j^{th} grid cell is normalized such that $\Delta_j = \lambda_e(r_{j-1})/\lambda_{e0}$, and thus $\Delta_1 = \lambda_e(r_0)/\lambda_{e0} = 1$. To determine the size of the second cell, first the position is determined (i.e., $r_1 = r_0 + \lambda_{e0}$) from which the Debye length $\lambda_e(r_1)$ is determined using the above density and temperature profiles given by equations (2) and (3), and then $\Delta_2 = \lambda_e(r_1)/\lambda_{e0}$. The next position r_2, is then determined from the equation $r_2 = r_1 + \lambda_e(r_1)$ and then the size of the third grid (Δ_3) is calculated, and the procedure continues on until the boundary at the high altitude end of the system is reached (which depends on the number of grid points used). Note that

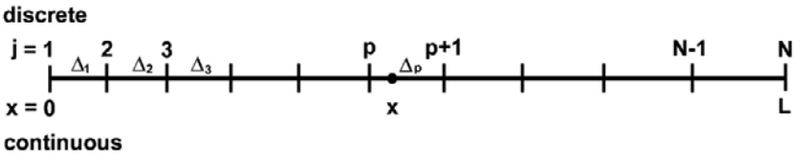


Fig. 2. An irregular grid system is shown whereby the grid size increases according to the local electron Debye length, which increases along the system axis (corresponding to an increasing altitude in the auroral zone with $x = r$). Discrete quantities are shown above the horizontal axis and continuous quantities are shown below the axis.

the size of each succeeding grid increases (i.e., $\Delta_{j+1} > \Delta_j$) and thus the grid size is larger at higher altitudes where the corresponding local Debye length is larger. This grid scheme is illustrated schematically in Fig. 2. The system length is $L = \sum \Delta_j$ (for $j = 1$ to N), which for the density and temperature profiles discussed above and about 50,000 grid points equals a distance of about 1 R_E. The grid size at the high altitude end of the system is about 10 times that at low altitudes in the ionosphere ($\Delta_N/\Delta_1 \approx 10$).

Once the variable sized grid system has been constructed according to the method illustrated in Fig. 2, the particles are then loaded in space. This is done by taking the number of particles per grid to be constant (typically 50 or 100 particles per grid) and then distributing the particles in space uniformly across each grid. For example, 50 particles per grid has been used in many of the auroral simulation runs. Thus groups of 50 particles are given a position equally spaced across each of the 50,000 grids (giving a total of 2,500,000 particles of one species, either ions or electrons). Since the grid size gets larger with system length, the density (particle number per grid length, i.e., $50/\Delta_j$) decreases. In the system described above the grid size increases exponentially with altitude and thus the density, using an equal number of particles per grid, decreases exponentially. The end effect is that of a dense plasma at low altitudes at the left end of the system that decreases in density exponentially giving way to a tenuous plasma at higher altitudes at the right end of the system.

When the particle positions are determined, the velocities are also determined using the temperature profile given by equation (3). This is done using a standard Maxwellian random number generator for each particle's velocity and multiplying it by a thermal velocity factor that depends on position. Using the temperature profile in equation (3) results in a plasma that has an increasing temperature along the length of the simulation system. An example of the simulation density and temperature profile described here will be given in Sect. 5 and shown in Fig. 4.

An equal number of particles of both species (ions and electrons) are loaded in configuration and velocity space as described above. For each electron, an ion is loaded at the same position (although obviously the charge state, mass and velocity are different), assuring exact charge neutrality initially within the system. Thus for the typical system parameters given above (50 particles per grid and 50,000 grids), a total of 5,000,000 particles are followed in the system

with exactly half of the particles being electrons and the other half being ions. Typical ion to electron mass ratios used for the runs here are 250 or 400.

The present version of the code is electrostatic and thus Poisson's equation is solved to find the electric field:

$$\nabla \cdot E = 4\pi\rho \tag{4}$$

The charge density is given by ρ. In the simulation the normalized electric field ($\tilde{E}$) is found at each grid point (j) using finite differencing, i.e.,

$$\tilde{E}_{j+1} = \tilde{E}_j + C\tilde{\rho}_j \tag{5}$$

Here C is a constant normalization factor given by the inverse of the number of particles per grid (e.g., 1/50). The normalized charge density at each grid point ($\tilde{\rho}_j$) is found by summing up the contribution of all the particles near that grid point, where a particle's contribution is a linear fraction of its total charge (= 1) distributed to the two nearest grid points based on the particle position between the two grid points. To solve (5), the electric field must be specified at one of the boundaries. In the auroral simulations, the electric field is taken to be zero in the magnetosphere such that $\tilde{E}_N = 0$, or equivalently at the right hand boundary of the system $E(x = L) = 0$.

One complication that arises when dealing with variable sized grids is finding the two grid points nearest the particle based on the particle position. In a uniform grid system all of the grid sizes are equal and can be set to 1 (i.e., $\Delta_j = \Delta = 1$). In this simple case the spatial position within the system is easily related to the grid number and finding the nearest grid points for a particular particle location is trivial. For example, in Fig. 2, if the grid spacing were taken to be uniform with $\Delta = 1$, a particle located at $x = 10.12$ would be nearest the grid points corresponding to $j = 11$ and $j = 12$. In a variable grid system this is no longer necessarily true. Thus a routine is needed to determine the particle position in a non-uniform grid system. Fortunately "locating" subroutines which use the bisection method to search an ordered table can be found in texts on numerical methods. The specific subroutine used here is called HUNT, (subroutine listing is given on page 91 in [33]), which is advantageous when searching a table many times that has nearly identical abscissas on consecutive searches.

Once the grid points nearest the particle positions have been found, a fundamental limit on the time step (Δt) for stability of PIC codes works to the benefit of finding the two nearest grid points for the particles on each succeeding time step. This fundamental limit, known as the Courant condition [7],[33], states that a particle should not move more than one grid length in a time step (i.e., $v\Delta t < \Delta$, where v is the particle velocity). Thus the nearest grid points for a particle can be found by keeping an auxiliary array that stores the nearest grid point at the initial time step, and then at each succeeding time step the nearest grid points will either be the same or change by at most one grid. This limits the search required to find the nearest grid points to either one or two steps and does not require a search over the entire grid system. If the nearest grid point

changes in a time step, the array value is updated for the next search on the next time step.

Another issue to deal with in a variable grid system is finite particle size. In most PIC codes, particles are given a finite size to reduce the noise caused by close Coulomb collisions and it has been shown that good results are obtained when the simulation particle is set equal to the grid size [7],[11]. In a variable grid system, since the grid size changes, the particle size will change as it moves around the system to match the local grid size. The grid size variation considered here is slow and monotonic ranging over thousands of grid points. Thus any particle traversing the system will see its size change slowly in space and time. The amount of charge on each particle however does not change and thus the contribution of a particle in the charge density to the two nearest grids will proportionally be the same regardless of location. Although the notion of a simulation particle that changes size depending on location might seem odd, keep in mind that the use of finite size particles is a numerical technique to keep the noise level low in PIC simulations.

We now look at two simulation runs where one run uses a variable sized grid system and the other uses a uniform sized grid system. The electron temperature time history for these two cases is shown in Fig. 3. The results for the two different runs clearly show that for a uniform grid system the cool plasma heats up due to finite grid size effects since the local Debye length does not match the grid size [7]. The variable grid system where the local grid size matches the Debye length clearly shows the elimination of the non-physical heating.

4 Non-periodic Boundary Conditions

Specific non-periodic boundary conditions are determined by the physical problem that is to be simulated. For example, to study the physics of shocks, beam injection at one end of the system and a reflection boundary condition at the other end of the system have been used in PIC codes (e.g., [3],[14]). Another set of PIC simulations applied to study active experiments near spacecraft has used beam injection along with various types of charged boundary conditions [5],[26],[41]. PIC simulations of the magnetotail that focussed on convection [30] and reconnection [31] also had to invoke a particular set of boundary conditions such as inflow from top and bottom and outflow or reflection at the sides, to examine the physical problem at hand. In these and other cases, the boundary conditions and system setup are determined by the problem being addressed.

For the auroral zone simulation considered here it is the satellite observations that determine how to treat the boundaries. For example, below the low altitude end of the simulation system (i.e., < 1000 km altitude) is the ionosphere. The ionosphere can be treated in one sense as a large reservoir of cold, dense plasma. Thus the particle boundary condition used here is that of a "cold reflector". What this means is that any particle that hits the low altitude boundary at $x = 0$ is reflected and given a return velocity based on a Maxwellian distribution with the temperature at $x = 0$. As discussed in Sect. 3, the temperature

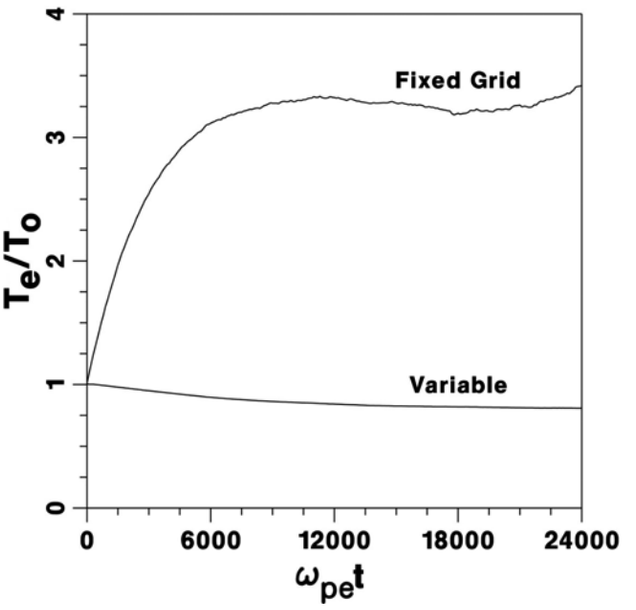


Fig. 3. This plot shows the electron temperature time history for two different simulation runs, one with a fixed grid size system and the other with a variable grid size system. For both simulation runs the plasma is initially loaded such that cool plasma is located at one end of the system and warmer plasma is located at the other end of the system. The case with variable grid points is set up such that the grid size approximately equals the local electron Debye length throughout the system, while the case with a fixed grid size sets the grid size equal to the Debye length of the warm plasma throughout the system.

follows a profile given by equation (3) and thus the reflected particle is given the temperature at $r = 1000$ km altitude. This condition is equivalent to replacing a precipitating particle with a cold ionospheric particle in its place. Thus if higher energy particles hit the low altitude boundary, energy can be lost from the system to the ionosphere. Note that this is a closed boundary condition in the sense that no particles are lost or removed from the system.

The high altitude end of the simulation system ($x = L$) is connected to the Earth's magnetotail plasma sheet, which in contrast to the ionosphere is a source of warm, energetic plasma with relatively low density. Since it is assumed that the magnetotail plasma sheet provides free energy to drive auroral dynamics, beam injection is included at the high altitude boundary. Specifically, new particles are introduced at $x = L$ with a negative beam drift velocity such that they move from right to left for the system setup shown in Fig. 2. A zero-current beam is injected meaning that both an ion and electron enter the system at the same time with the same net drift velocity. Any particles from inside the system that hit the high altitude boundary at $x = L$ are reflected back into the

system. Note that for these boundary conditions (cold reflection at $x = 0$ and standard reflection at $x = L$) and the injection of a zero-current beam at $x = L$, there are always an equal (but increasing) number of electrons and ions in the system throughout the simulation run. The beam injection velocities for ions and electrons are based on satellite observations and a number of different beam speeds have been tried. Recent Polar satellite observations indicate that large amplitude Alfven wave Poynting fluxes flow from the magnetotail into the auroral zone [43] and represent another free energy source for auroral processes. Future simulations will consider plasma wave injection at the high altitude boundary.

Another set of boundary conditions now being considered for the auroral PIC simulations are to include a net current across the system. This is dictated by satellite observations that show field aligned currents flowing either towards or away from the Earth in the auroral zone (e.g., [8]). The boundary conditions discussed above are such that charge neutrality is strictly conserved over the entire system and there is no net gain or loss of charge. This includes the injected beam particles in the sense that ion and electron beam speeds are set equal and there is a net zero current inflow (or outflow) across the system boundary. This condition needs to be relaxed in light of satellite observations in the auroral zone and there are various ways of doing this. One way is to inject ions and electrons with a different net drift from the magnetotail boundary. This has been done in two ways, one where an equal number of ions and electrons are input at the boundary but they are given a different drift speed. Although this allows the current to be non-zero within the system, there is no build up of charge and thus no flow of current across the boundary. The result is that while current-driven instabilities can be examined, fields due to an excess of charge and the redistribution of the plasma in the system is not included. Runs have also been performed whereby more electrons flow into the system creating a field-aligned current and a resulting charge imbalance. The difficulty comes in allowing the charge buildup to be alleviated through the loss of (certain) charged particles at either boundary providing a field-aligned current through the entire system from magnetotail to ionosphere. Various possibilities are presently being considered to handle this situation, including the inclusion of an electric potential drop being imposed across the system based on a current-voltage relationship appropriate to the auroral zone [21].

5 Example of Irregular Grids and Non-periodic Boundaries: The Auroral Zone

We now consider in detail the PIC simulation code described in the previous section applied to the auroral zone. The auroral zone is a very active region and is characterized by plasma acceleration, wave-particle interactions and quasi-static parallel electric fields. Questions that remain to be answered in the auroral zone include how quasi-static parallel electric fields are formed, what is the role of wave-particle interactions and ultimately what is the free energy source for the precipitating particles. To examine some of these issues the auroral PIC

simulation has been employed. To illustrate how the code works some of the results will be presented here (see [34] for more details).

A mechanism that can lead to quasi-static parallel electric fields in the auroral zone is the magnetic mirroring of ion and electron distribution functions at different altitudes as a result of an anisotropy in the plasma sheet [2],[29]. Many static models of the steady-state auroral electrostatic potential in the primary current region are based on this premise [9],[17],[23],[39],[40]. A particular type of anisotropy that can generate upward directed quasi-static electric fields are earthward streaming ion beams that originate from the magnetotail [17],[18]. This case is considered here.

To study the self-consistent formation of parallel electric fields in the auroral zone, the PIC simulation code with variable grid spacing and non-periodic boundary conditions as described in the previous sections is used. For the cases discussed here only electrostatic processes are considered and the code is one-dimensional along an auroral magnetic field line. An additional feature of the code not already discussed is the presence of the mirror force. This is important because closer to the Earth the magnetic field strength increases due to the Earth's dipole and for charged particles the mirror force can play an important role over distances of thousands of kilometers. The equation of motion used to push the ions and electrons in the system for electrostatic waves in the presence of the mirror force is given by:

$$ma = qE - \mu\nabla_{\parallel}B - mg \tag{6}$$

The electric field is given by E, the particle mass is given by m, μ is the magnetic moment, g represents gravity and $\nabla_{\parallel}B$ is the parallel gradient of the terrestrial magnetic field (i.e., the mirror force term). A background of cold dense plasma that represents the ionosphere is loaded at the low altitude end of the simulation box as discussed in Sect. 3 and the system is driven by warm tenuous anisotropic ion and electron particle distribution functions injected at the high altitude end of the system. The magnetic field goes as a dipole along the magnetic field direction, which is the same as the simulation axis, such that $B(x)$ goes as $1/x^3$. The magnetic moment of each particle is calculated at $t = 0$ using $\mu = W_{\perp}/B(x)$, based on the particle's initial x position ($W_{\perp}$ is the perpendicular energy). Adiabatic invariance is assumed such that μ is constant for every particle throughout the simulation run. A schematic of the simulation system and its location in the auroral zone was shown in Fig. 1.

The run presented here will consider a beam of ions and electrons from the high altitude magnetotail end of the system. This simulates earthward streaming plasma in the Earth's plasma sheet boundary layer (PSBL) and represents a free energy source for the auroral zone. Ion and electron beams are a common feature of the PSBL [13],[15],[16],[28] and are a possible driver of auroral processes [18],[25],[38].

Figure 4 shows the initial phase space for the protons in the top panel, the electrons in the middle panel and the density and magnetic field profiles in the bottom panel. The ion phase space (top panel of Fig. 4) shows the incoming

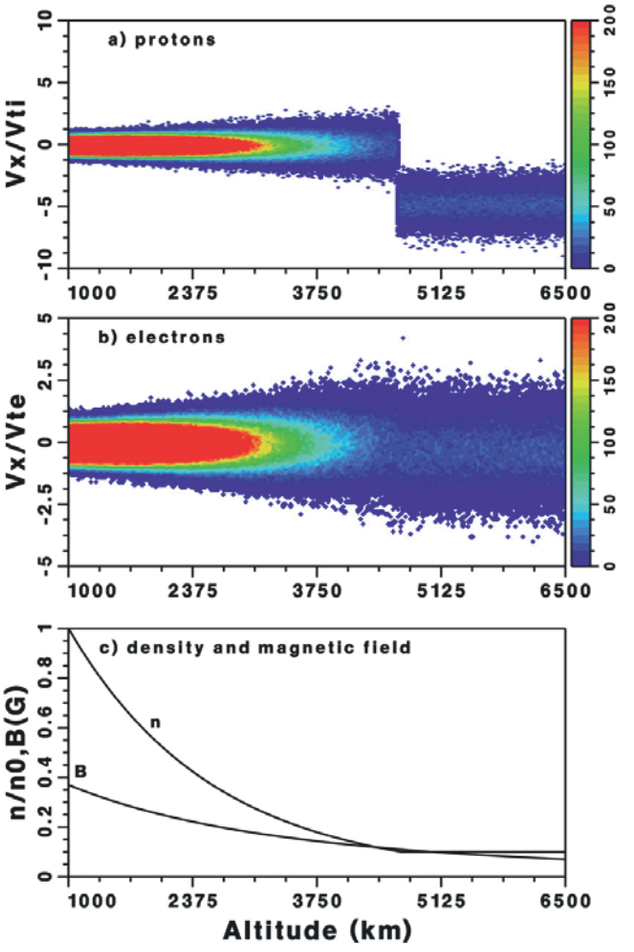


Fig. 4. This plot shows the intial phase space of protons and electrons in the top two panels and the initial density and magnetic field profile in the bottom panel. These quantities are shown along the simulation (horizontal) axis, which is equivalent to altitude. The phase space diagrams show velocity normalized to their respective thermal velocities along the vertical axis. The color code is such that red represents regions of higher phase space density and blue corresponds to lower phase space densities.

ion beam flowing from the right into the ionospheric background. Here there is no net current thus the electrons and ions have the same drift speed ($U_i = U_e$), however, the velocities in Fig. 4 are shown relative to the species thermal velocity. Since $U_e/v_{te} < 1$, the beam is less apparent for electrons. Another item to notice in Fig. 4 is that the plasma density is high and the plasma is cooler at lower altitudes at the left end of the simulation system.

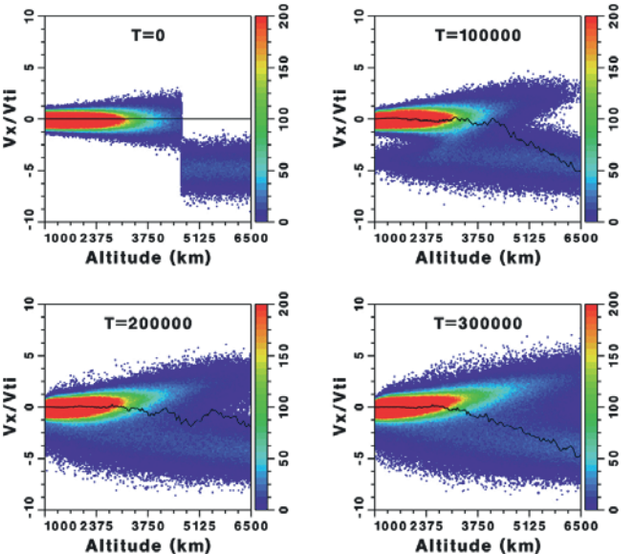


Fig. 5. Snapshots of the proton phase space at four different times are shown. The initial configuration at the beginning of the run is shown in the upper left panel, and the bottom right panel shows the results near the end of the simulation run. The black line running across the system is the electrostatic potential in kilovolts at that time.

Figure 5 shows the ion phase space and the average electrostatic potential at four different times during the run. As time progresses the ion beam in Fig. 5 can be seen streaming earthward (from right to left) and interacts with both the background ionospheric ions and with an increasing magnetic field (mirror force) that tends to repel the incoming ions. Over time a potential drop (shown by the black solid line) forms due to a net charge separation between the inflowing ions and electrons [2],[29]. This charge separation is caused by the mirror force and the anisotropy (beam) of the injected plasma [34],[37].

The strength of the potential drop is a function of the injected drift speed of the ions [38]. This is borne out in Fig. 6, which shows the average (over the course of the run) electrostatic potential across the system for three runs with different inflow beam speeds. It can seen in Fig. 6 that the higher the drift speed the larger the potential (given in kilovolts). This makes intuitive sense since a larger beam speed implies more free energy being pumped into the system. Physically the larger drift speed means a larger charge separation between the ions and the electrons, and hence a larger potential drop.

The last result to be discussed from this set of PIC simulation runs is the generation of plasma waves within the simulation system. Because beams are present and there is thermal mixing of plasmas, it is expected that instabilities will result. This is illustrated in Fig. 7, which shows color-coded wave power with frequency versus the system length. One result immediately apparent in Fig. 7 is the intense wave activity excited at low frequencies in the mid-altitude

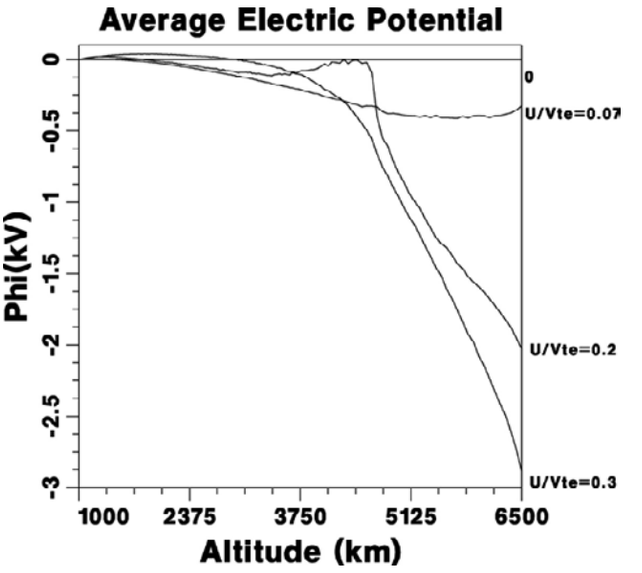


Fig. 6. This plot shows the average electrostatic potential across the simulation system for three different simulation runs with different inflow beam speeds (shown at the right). The potential is time averaged over the entire run and is shown in kilovolts versus altitude (in km) along the simulation system.

range ($\sim$ 4000 km) where the injected beam plasma from the magnetotail (at the right) interacts with the cool ionospheric plasma (at the left). These are likely to be ion-ion acoustic waves excited by the injected ion beam [34]. Another interesting feature in Fig. 7 is the presence of local electron plasma oscillations, which can be seen by the enhancements at higher frequency that are highest at low altitudes (1000 km) and tend to decrease in frequency with altitude. This decreasing frequency profile is simply due to the decrease in ionospheric density that was initiated in the system. For more details concerning this simulation code and results discussing the implications for the auroral zone, the reader is referred to [34],[35],[36],[37].

6 Conclusions

Particle in cell simulations with variable sized grid cells and non-periodic boundary conditions have been discussed here with results presented for a model of the auroral zone. As the spatial domain in PIC simulation systems grow in size covering different regions in space plasmas, variable grid configurations will be become a necessary element to capture the essential physics. The PIC simulation presented here to model the auroral zone is relatively simple in the sense that it is one dimensional and electrostatic, yet it yielded valuable results pertaining to the formation of quasi-static parallel potential drops and the effects of wave-

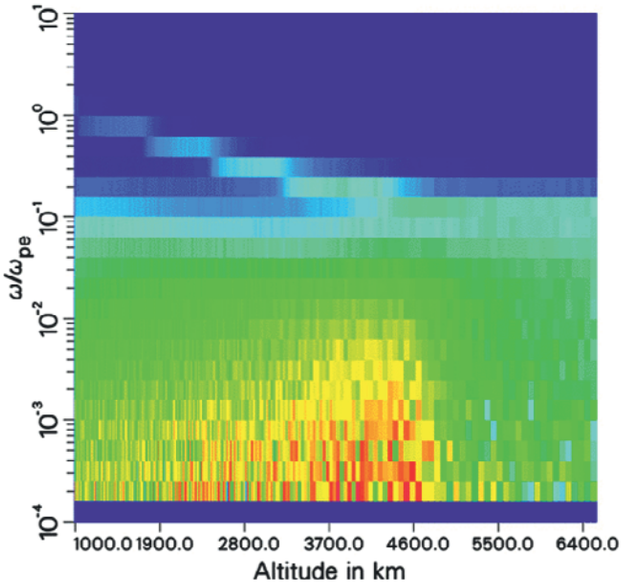


Fig. 7. This plot shows an electric field wave spectogram with frequency plotted versus system length. The plot is color-coded in wave power with red representing higher electric field intensity and blue lower electric field intensity. The frequency (vertical axis) is normalized to the electron plasma frequency at 1000 km.

particle interactions. Even such a simplified version of the code (one dimension, electrostatic), however, can be computationally demanding. For example, the run presented in Sect. 5, which used about 50,000 grid points, 5,000,000 total particles and was run about 1,500,000 time steps, required a total of nearly 1300 hours of computing time on a Cray T90.

The physics of the auroral zone ultimately demands higher dimensional, electromagnetic simulations, which will be the next step. In general this will be the case in other regions of space as well. Fortunately computing speed and available memory have increased dramatically in the last few years and such systems are now feasible from a computational point of view. It seems clear that, at least for the auroral zone, an effective two dimensional system will have to be non-Cartesian to allow for the needed resolution and variable grid spacing required to model the plasma. This will surely present a new set of challenges for the grid sizing and the boundary conditions to be tackled in the future.

Acknowledgements

The author wishes to thank M. Ashour-Abdalla and R. Richard for their considerable help over the course of this entire project. This work was supported by NASA Guest Investigator Grant NAG5-10473, NASA ISTP Grant NAG5-11704 and NSF Grant INT-0010111. Computing resources were provided by the

National Partnership for Advanced Computing Infrastructure (NPACI) and in particular the San Diego Supercomputing Center (SDSC).

References

1. S.I. Akasofu, E.W. Hones, Jr., S.J. Bame, J.R. Asbridge, A.T.Y. Lui: J. Geophys. Res. **78**, 7257 (1973)
2. H. Alfven, C.G. Falthammer: *Cosmical Electrodynamics* (Clarendon, Oxford 1963)
3. C.H. Aldrich: Space Sci. Rev. **42**, 131 (1985)
4. V. Angelopoulos, T.D. Phan, D.E. Larson, F.S. Mozer, R.P. Lin, K. Tsuruda, H. Hayakawa, T. Mukai, S. Kokubun, T. Yamamoto, D.J. Williams, R.W. McEntire, R.P. Lepping, G.K. Parks, M. Brittnacher, G. Germany, J. Spann, H.J. Singer, K. Yumoto: Geophys. Res. Lett. **24**, 2271 (1997)
5. M. Ashour-Abdalla, H. Okuda: Adv. Space Res. **8**, (1)137 (1988)
6. W. Baumjohann, G. Paschmann, H. Luhr: J. Geophys. Res. **95**, 3801 (1990)
7. C.K. Birdsall, A.B. Langdon: *Plasma Physics Via Computer Simulations* (McGraw-Hill, New York 1985)
8. C.W. Carlson, J.P. McFadden, R.E. Ergun, M. Temerin, W. Peria, F.S. Mozer, D.M. Klumpar, E.G. Shelley, W.K. Peterson, E. Moebius, R. Elphic, R. Strangeway, C. Catell, R. Pfaff: Geophys. Res. Lett. **25**, 2017 (1998)
9. Y.T. Chiu, J.M. Cornwall: J. Geophys. Res. **85**, 543 (1980)
10. J.M. Dawson: Phys. Fluids **5**, 445 (1962)
11. J.M. Dawson: Rev. Modern Phys. **55**, 403 (1983)
12. T.E. Eastman, E.W. Hones, Jr., S.J. Bame, J.R. Asbridge: Geophys. Res. Lett. **3**, 685 (1976)
13. T.E. Eastman, L.A. Frank, W.K. Peterson, W. Lennartsson: J. Geophys. Res. **89**, 1553 (1984)
14. D.W. Forslund, K.B. Quest, J.U. Brackbill, K. Lee: J. Geophys. Res. **89**, 2142 (1984)
15. L.A. Frank, K.L. Ackerson, R.P. Lepping: J. Geophys. Res. **89**, 5859 (1976)
16. E.W. Hones, Jr., J.R. Asbridge, S.J. Bame, M.D. Montgomery, S. Singer, S.-I. Akasofu: J. Geophys. Res. **77**, 5503 (1972)
17. J.R. Kan: J. Geophys. Res. **80**, 2089 (1975)
18. J.R. Kan, S.I. Akasofu: J. Geophys. Res. **81**, 5123 (1976)
19. C.A. Kletzing, R.B. Torbert: J. Geophys. Res. **99**, 2159 (1994)
20. A.J. Klimas: J. Geophys. Res. **88**, 9081 (1983)
21. S. Knight: Planet. Space Sci. **21**, 741 (1973)
22. J.N. LeBoeuf, T. Tajima, C.F. Kennel, J.M Dawson: Geophys. Res. Lett. **5**, 609 (1978)
23. J. Lemaire, M. Scherer: Planet. Space Sci. **22**, 1485 (1974)
24. A.T.Y. Lui: 'Micro/Mesoscale coupling in the magnetotail current sheet: Observations'. In: *Cross-Scale Coupling in Space Plasmas, Geophys. Mono. Ser. 93*, ed. by J.L. Horwitz, N. Singh, J.L. Burch (American Geophysical Union, Washington, D.C. 1995) pp. 261
25. L.R. Lyons: 'Discrete auroras and magnetotail processes'. In: *Auroral Physics*, ed. by C.-I. Meng, M.J. Rycroft, L.A. Frank (Cambridge University Press, New York 1991) pp. 195
26. H. Okuda, R. Horton, M. Ono, M. Ashour-Abdalla: Phys. Fluids **30**, 200 (1987)

27. G. Paschmann, G. Haerendel, N. Sckopke, H. Rosenbauer, P.C. Hedgecock: J. Geophys. Res. **81**, 2883 (1976)
28. G.K. Parks, M. McCarthy, R.J. Fitzenreiter, J. Etcheto, K.A. Anderson, R.R. Anderson, T.E. Eastman, L.A. Frank, D.A. Gurnett, C. Huang, R.P. Lin, A.T.Y. Lui, K.W. Ogilvie, A. Pedersen, H. Reme, D.J. Williams: J. Geophys. Res. **89**, 8885 (1984)
29. H. Persson: Phys. Fluids **6**, 1756 (1963)
30. P.L. Pritchett, F.V. Coroniti: J. Geophys. Res. **100**, 23551 (1995)
31. P.L. Pritchett: J. Geophys. Res. **106**, 3783 (2001)
32. D. Potter: *Computational Physics* (Wiley, New York 1973)
33. W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling: *Numerical Recipes: The Art of Scientific Computing (FORTAN Version)* (Cambridge University Press, Cambridge 1989)
34. D. Schriver: J. Geophys. Res. **104**, 14655 (1999)
35. D. Schriver, M. Ashour-Abdalla: Geophys. Res. Lett. **20**, 475 (1993)
36. D. Schriver, M. Ashour-Abdalla: 'Particle acceleration in the auroral zone: Comparison of simulations with observations'. In: *Physics of Space Plasmas (1998), No. 15*, ed. by T. Chang, J.R. Jasperse (MIT Center for Theoretical Geo/Cosmo Plasma Physics, Cambridge 1998) pp. 301
37. D. Schriver, M. Ashour-Abdalla, R.L. Richard: Phys. Chem. Earth (C) **26**, 65 (2001)
38. Y. Serizawa, T. Sato: Geophys. Res. Lett. **11**, 595 (1984)
39. D.W. Swift: J. Geophys. Res. **80**, 2096 (1975)
40. E.C. Whipple: J. Geophys. Res. **82**, 1525 (1977)
41. R.M. Winglee, P.L. Pritchett: J. Geophys. Res. **92**, 6114 (1987)
42. D. Winske, M.M. Leroy: 'Hybrid simulation techniques applied to the Earth's bow shock', In: *Computer Simulations of Space Plasmas*, ed. by H. Matsumoto T. Sato (D. Reidel Terra, Hingham, Mass. 1984)
43. J.R. Wygant, A. Keiling, C.A. Cattell, M. Johnson, R.L. Lysak, M. Temerin, F.S. Mozer, C.A. Kletzing, J.D. Scudder, W. Peterson, C.T. Russell, G. Parks, M. Brittnacher, G. Germany, J. Spann: J. Geophys. Res. **105**, 18675 (2000)
44. M. Ashour-Abdalla, D.A. Dutton (Eds.): *Space Plasma Simulations* (D. Reidel, Dordrecht 1985)
45. B. Lembege, J.W. Eastwoood (Eds.): *Numerical Simulation of Space Plasmas* (Elsevier, Amsterdam 1988)

Low Noise Electrostatic and Electromagnetic Delta-f Particle-in-Cell Simulation of Plasmas

Richard D. Sydora

Department of Physics, University of Alberta, Edmonton, Canada T6G 2J1

Abstract. This paper presents a method for performing efficient and accurate particle simulations of weakly growing instabilities found in both laboratory and space plasmas. This low noise technique is based on a particle-in-cell (PIC) formulation of the equations for the perturbed plasma charged particle distribution function. This technique substantially reduces the number of particles one needs for discrete sampling since only the perturbed distribution is represented by particles. In this work results of simulations are presented from several classes of electrostatic and electromagnetic instabilities driven by velocity space nonuniformity. Simulation results from studies of linear and nonlinear Landau damping are also presented.

1 Introduction

The dynamics of kinetic microinstabilities in space plasmas, particularly in the weak turbulence regime (linear instability growth, $\gamma < \omega$, the real frequency), are very difficult to investigate using conventional particle-in-cell (PIC) plasma simulation models [1,11]. This is due to the poor signal-to-noise ratio when attempting to resolve the entire distribution funtion with discrete particles. Using the conventional PIC method, random statistical fluctuations in the number of particles per cell causes field fluctuations. It is very difficult to reduce this noise by increasing the number of particles, N, since statistical noise only decreases as $N^{-1/2}$ for random loading and approximately N^{-1} for the quiet start methods which are a more uniform particle loading in phase space [1,24]. For many problems of interest the physical fluctuations of the kinetic distribution function, δf, are often much smaller than the total distribution, f, and can be dominated by the noise. The δf algorithm alleviates this problem since only the perturbed part of the distribution is represented by particles. This leads to orders of magnitude reduction in the noise level, particularly when f_o can be chosen such that $\delta f/f_o < 1$. This is because statistical fluctuations of the total f become $\delta N \delta f/N$ where δN is the statistical fluctuation in the number of particles per cell.

The δf-PIC method was originally developed in the electrostatic 1D Vlasov-Possion system [13], building on previous work, and extended to the drift-kinetic and gyrokinetic low frequency plasma descriptions [7,20,12,5,23]. Further generalizations to the fully electromagnetic PIC simulation, including relativistic effects, have recently been made [24]. Using this technique it is possible to more clearly identify the role of wave-wave-wave and wave-wave-particle interactions in the saturation dynamics of kinetic microinstabilities driven by free energy

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 109–124, 2003.

sources in the distribution function and temperature and density spatial inhomogeneity.

In this paper the detailed algorithm for the complete electromagnetic system is presented along with several case studies of electrostatic and electromagnetic instabilities. These examples were chosen because they are difficult to simulate using the conventional PIC method. The standard examples of linear and nonlinear Landau damping, weak beam instability, and thermal anisotropy-driven electromagnetic instabilities, are found in many introductory graduate plasma physics textbooks [8,14,3,10,17]. These are good supplements to the simulation examples presented here. These examples are also of contemporary interest. For instance, the weak beam instability nonlinear evolution has recently been connected to the generation of electrostatic solitary waves in the magnetotail as observed by the GEOTAIL spacecraft [18]. Studies of the beam instabilities have been made using conventional PIC methods with approximately 10 million particles in 1D and 30 million in 2D [16].

2 δF Particle Model

2.1 Electromagnetic Model Equations

In this section the basic equations for the electromagnetic δf-PIC algorithm are presented. A discussion of total-f electromagnetic simulation models, including examples, is given in the chapter by Pritchett [21]. We assume a collisionless plasma and the evolution of the distribution function for a particular species is governed by the Vlasov equation in continuity form

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial \boldsymbol{z}} \cdot (\dot{\boldsymbol{z}} f) = 0 \tag{1}$$

where the phase space coordinate is represented by the general vector $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{v})$. By splitting the distribution function into a background and perturbation parts, $f(\boldsymbol{z}, t) = f_o(\boldsymbol{z}) + \delta f(\boldsymbol{z}, t)$, we can write the equation for the evolution of δf

$$\frac{\partial \delta f}{\partial t} + \frac{\partial}{\partial \boldsymbol{z}} \cdot (\dot{\boldsymbol{z}} \delta f) = -\dot{\boldsymbol{z}_1} \cdot \frac{\partial f_o}{\partial \boldsymbol{z}} \tag{2}$$

where $\dot{\boldsymbol{z}_1} = (0, (q/m)[\boldsymbol{E}_1 + \frac{\boldsymbol{v} \times \boldsymbol{B}_1}{c}])$. The characteristics, or particles, follow the full nonlinear trajectories, $\boldsymbol{z} = \boldsymbol{z}_o + \boldsymbol{z}_1$, and $\dot{\boldsymbol{z}_o} = (\boldsymbol{v}, (q/m)[\boldsymbol{E}_o + \frac{\boldsymbol{v} \times \boldsymbol{B}_o}{c}])$.

The perturbed distribution,δf, is represented by

$$\delta f(\boldsymbol{z}, t) = \sum_i w_i \delta(\boldsymbol{z} - \boldsymbol{z}_i) \tag{3}$$

where a particle weight is defined as

$$w_i = \frac{\delta f}{g} \tag{4}$$

g is a numerically loaded particle distribution or marker distribution expressed as

$$g(\boldsymbol{z},t) = \sum_i \delta(\boldsymbol{z} - \boldsymbol{z}_i) \tag{5}$$

and is arbitrary. The equation of motion for the weights, w_i, is obtained by substituting these representations for δf and g into Eq.(2) and we obtain

$$\frac{dw_i}{dt} = -\dot{\boldsymbol{z}}_1 \cdot \frac{1}{g(\boldsymbol{z},t)} \frac{\partial f_o}{\partial \boldsymbol{z}}\bigg|_{\boldsymbol{z}} \tag{6}$$

Since both f and g satisfy df/dt=0 and dg/dt=0 we can express the source term on the right hand side of Eq.(6) as

$$-\tfrac{f_o}{g(\boldsymbol{z},t=0)}\dot{\boldsymbol{z}}_1 \cdot \tfrac{1}{f_o(\boldsymbol{z})}\tfrac{\partial f_o}{\partial \boldsymbol{z}} = -\left(\tfrac{f(0)}{g(0)} - w_i\right)\dot{\boldsymbol{z}}_1 \cdot \tfrac{1}{f_o(\boldsymbol{z})}\tfrac{\partial f_o}{\partial \boldsymbol{z}} \tag{7}$$

In summary, the final equations which we solve in the electromagnetic δf-PIC model consists of the nonlinear trajectory equations, the particle weight equations and Maxwell's equations for the the electromagnetic fields [24].

$$\begin{gathered}
\tfrac{dw_i}{dt} = -\left(\tfrac{f(t=0)}{g(t=0)} - w_i\right)\dot{\boldsymbol{z}}_1 \cdot \tfrac{1}{f_o(\boldsymbol{z})}\tfrac{\partial f_o}{\partial \boldsymbol{z}} \\
\dot{\boldsymbol{z}_o} = (\boldsymbol{v}, (q/m)[\boldsymbol{E}_o + \tfrac{\boldsymbol{v}\times\boldsymbol{B}_o}{c}]) \\
\dot{\boldsymbol{z}_1} = (0, (q/m)[\boldsymbol{E}_1 + \tfrac{\boldsymbol{v}\times\boldsymbol{B}_1}{c}]) \\
\boldsymbol{\nabla}\times\boldsymbol{E}_1 = -\tfrac{1}{c}\tfrac{\partial \boldsymbol{B}_1}{\partial t} \\
\boldsymbol{\nabla}\times\boldsymbol{B}_1 = \tfrac{1}{c}\tfrac{\partial \boldsymbol{E}_1}{\partial t} + \tfrac{4\pi}{c}\textstyle\sum_s\sum_{i=1}^N q_i w_i \boldsymbol{v}_i S(\boldsymbol{x}-\boldsymbol{x}_i) \\
\boldsymbol{\nabla}\cdot\boldsymbol{E}_1 = 4\pi\textstyle\sum_s\sum_{i=1}^N q_i w_i S(\boldsymbol{x}-\boldsymbol{x}_i)
\end{gathered} \tag{8}$$

2.2 Finite Difference Equations

The integration of Eqs.(8) is briefly discussed in this section. The time-centered integration of Maxwell's equations and the particle equations of motion are presented in the chapter by Pritchett [21] and some of the steps are repeated here for completeness.

We consider explicit time integration schemes for the temporal differencing and Fourier representation for the spatial derivatives. The FFT algorithm is efficiently used to transform between real and k-space. Implicit time differencing can easily also formulated within the δf-PIC framework and allow for the study of low frequency fluctuations [15].

To proceed with the solution of the field equations we use a spectral decomposition for the electric and magnetic fields, separating Maxwell's equations into longitudinal and transverse parts (($\boldsymbol{E}_L(\boldsymbol{k}) = \boldsymbol{k}(\boldsymbol{k}\cdot\boldsymbol{E})/k^2$, $\boldsymbol{E}_T(\boldsymbol{k}) =$

$\boldsymbol{E}(\boldsymbol{k}) - \boldsymbol{E}_L(\boldsymbol{k})$). The field equations in Fourier space are [6]

$$\begin{array}{c} \frac{\partial \boldsymbol{E}_T(\boldsymbol{k},t)}{\partial t} = ic\boldsymbol{k} \times \boldsymbol{B}(\boldsymbol{k},t) - 4\pi \boldsymbol{J}_T(\boldsymbol{k},t) \\ \frac{\partial \boldsymbol{B}(\boldsymbol{k},t)}{\partial t} = -ic\boldsymbol{k} \times \boldsymbol{E}_T(\boldsymbol{k},t) \\ \boldsymbol{E}_L = -4\pi i \boldsymbol{k} \rho(\boldsymbol{k},t)/k^2 \end{array} \tag{9}$$

and the time-centered Leapfrog finite-difference scheme gives

$$\begin{array}{c} \boldsymbol{E}_T^{n+1/2}(\boldsymbol{k}) = \boldsymbol{E}_T^{n-1/2}(\boldsymbol{k}) + ic\Delta t \boldsymbol{k} \times \boldsymbol{B}^n(\boldsymbol{k}) - 4\pi \Delta t \boldsymbol{J}_T^n(\boldsymbol{k}) \\ \boldsymbol{B}^{n+1}(\boldsymbol{k}) = \boldsymbol{B}^n(\boldsymbol{k}) - ic\Delta t \boldsymbol{k} \times \boldsymbol{E}_T^{n+1/2}(\boldsymbol{k}) \end{array} \tag{10}$$

The particle equations are integrated also using the Leapfrog method in finite-difference form

$$\begin{array}{c} \boldsymbol{x}_i^{n+1/2} = \boldsymbol{x}_i^{n-1/2} + \Delta t \boldsymbol{v}_i^n \\ \boldsymbol{v}_i^{n+1} = \boldsymbol{v}_i^n + \frac{q_i \Delta t}{m_i} [\boldsymbol{E}_i^{n+1/2} + \frac{(\boldsymbol{v}_i^{n+1} + \boldsymbol{v}_i^n) \times \boldsymbol{B}_i^{n+1/2}}{2c}] \end{array} \tag{11}$$

Integration of the weight evolution equation can be accomplished using the centered difference form

$$w_i^{n+1} = w_i^n - \Delta t \left(\frac{f(0)}{g(0)} - w_i^n \right) \dot{\boldsymbol{z}}_1^{n+1/2} \cdot \left(\frac{1}{f_o(\boldsymbol{z})} \frac{\partial f_o}{\partial \boldsymbol{z}} \right)^{n+1/2} \tag{12}$$

The weighted particles, or the δf portion of the distribution, is interpolated onto a grid for construction of the mesh charge and current densities. This is represented by the interpolation shape factor, S, in Eq.(8). Typically a second order or quadratic spline method is used.

2.3 Computational Cycle – Electrostatic Model

The detailed steps in the computational cycle for the electrostatic δf-PIC model are presented in this section. The main difference between this model and the conventional PIC is the weight evolution equation which must be adapted to the time-centered Leapfrog integration. The staggered particle position and velocity means that the weight time level must be adjusted to both.

Initially: $\boldsymbol{x}_i^{n-1/2}, \boldsymbol{v}_i^n, w_i^n, w_i^{n-1}$ are given.
Step 1: Advance $\boldsymbol{x}_i^{n-1/2}$ to $\boldsymbol{x}_i^{n+1/2}$ using $\boldsymbol{x}_i^{n+1/2} = \boldsymbol{x}_i^{n-1/2} + \Delta t \boldsymbol{v}_i^n$.
Determine $w_i^{n+1/2}$ using $w_i^{n+1/2} = (3/2)w_i^n - (1/2)w_i^{n-1}$.
Step 2: Accumulate the perturbed charge density on the grid using $\rho^{n+1/2}(\boldsymbol{x}) = \sum_{i=1}^N q_i w_i^{n+1/2} S(\boldsymbol{x} - \boldsymbol{x}_i^{n+1/2})$.
Step 3: Transform $\rho^{n+1/2}$ to k-space giving $\rho^{n+1/2}(\boldsymbol{k})$.
Solve for $\boldsymbol{E}^{n+1/2}(\boldsymbol{k})$ using $\rho^{n+1/2}(\boldsymbol{k})$ from Eq.(9).
Step 4: Transform $\boldsymbol{E}^{n+1/2}(\boldsymbol{k})$ to real space.
Step 5: Advance the velocities to $\boldsymbol{v}_i^{n+1}$ using the interpolated electric field at the particle and $\boldsymbol{v}_i^{n+1} = \boldsymbol{v}_i^n + \Delta t \boldsymbol{E}_i^{n+1/2}(\boldsymbol{x}^{n+1/2})$.
Step 6: Advance the weights using Eq.(12) and $v_i^{n+1/2} = (v_i^{n+1} + v_i^n)/2$ is used in the source term.
Step 7: Return to Step 1.

2.4 Computational Cycle – Electromagnetic Model

We now summarize the steps involved in advancing the electromagnetic fields and particles for one time step.

Initially: $\boldsymbol{x}_i^{n-1/2}, \boldsymbol{v}_i^n, w_i^n, w_i^{n-1}, \boldsymbol{E}_T^{n-1/2}, \boldsymbol{B}^n$ are given.
Step 1: Advance $\boldsymbol{x}_i^{n-1/2}$ to $\boldsymbol{x}_i^n$ using $\boldsymbol{x}_i^n = \boldsymbol{x}_i^{n-1/2} + (\Delta t)\boldsymbol{v}_i^n$.
Step 2: Accumulate the perturbed current density on the grid using $\boldsymbol{J}^n(\boldsymbol{x}) = \sum_{i=1}^N q_i \boldsymbol{v}^n w_i^n S(\boldsymbol{x} - \boldsymbol{x}_i^n)$ and transform $\boldsymbol{J}^n(\boldsymbol{x})$ to k-space giving $\boldsymbol{J}^n(\boldsymbol{k})$.
Step 3: Advance $(\boldsymbol{E}_T^{n-1/2}(\boldsymbol{k}), \boldsymbol{B}^n(\boldsymbol{k}))$ to $(\boldsymbol{E}_T^{n+1/2}(\boldsymbol{k}), \boldsymbol{B}^{n+1}(\boldsymbol{k}))$ from Eq.(10).
Step 4: Advance the positions using $\boldsymbol{x}_i^{n+1/2} = \boldsymbol{x}_i^{n-1/2} + \Delta t \boldsymbol{v}_i^n$. Determine $w_i^{n+1/2}$ using $w_i^{n+1/2} = (3/2)w_i^n - (1/2)w_i^{n-1}$.
Step 5: Accumulate the perturbed charge density on the grid using $\rho^{n+1/2}(\boldsymbol{x}) = \sum_{i=1}^N q_i w_i^{n+1/2} S(\boldsymbol{x} - \boldsymbol{x}_i^{n+1/2})$.
Step 6: Transform $\rho^{n+1/2}$ to k-space giving $\rho^{n+1/2}(\boldsymbol{k})$.
Solve for $\boldsymbol{E}_L^{n+1/2}(\boldsymbol{k})$ using $\rho^{n+1/2}(\boldsymbol{k})$ from Eq.(9).
Step 7: Transform $(\boldsymbol{E}_L^{n+1/2}(\boldsymbol{k}), \boldsymbol{E}_T^{n+1/2}(\boldsymbol{k}), \boldsymbol{B}^{n+1/2}(\boldsymbol{k}))$ to real space. Obtain $\boldsymbol{B}^{n+1/2} = (\boldsymbol{B}^{n+1} + \boldsymbol{B}^n)/2$ on the grid.
Step 8: Advance the velocities to $\boldsymbol{v}_i^{n+1}$ using the interpolated electric and magnetic fields at the particle as given in Eq.(11). Advance the weights using Eq.(12) and $v_i^{n+1/2} = (v_i^{n+1} + v_i^n)/2$ in the source term.

The evolution equation for the perturbed distribution also contains important conservation properties such as total number density

$$\sum_{i=1}^{N} w_i(t) = 0 \tag{13}$$

which must be monitored during the simulation. Other global quantities which can be output include the kinetic energy, $\sum_s \sum_{i=1}^N (m_s/2) v_i^2 w_i(t)$, and the total field energy, $\frac{1}{8\pi} \int d^3\boldsymbol{x}(E^2 + B^2)$.

3 δF-PIC Simulation Model Tests

In this section the results from simulations using the electrostatic and electromagnetic δf-PIC model are presented. The case studies are chosen from situations where conventional PIC methods are inaccurate or give results which are difficult to interpret. These include linear and nonlinear wave damping, weakly growing instabilities and evaluation of marginal stability.

3.1 Linear Landau Damping

As a first test case we consider the standard problem of linear Landau damping in a uniform unmagnetized plasma. The analytical damping rate can be obtained

from the electrostatic dispersion relation $\epsilon(k, \omega_r + i\gamma) = 0$, where the dielectric function is defined by [14]

$$\epsilon(k, \omega_r + i\gamma) = 1 + \sum_{j=species} \frac{\omega_{pj}^2}{k^2} \int d^3\boldsymbol{v} \frac{\boldsymbol{k} \cdot \frac{\partial f_j}{\partial \boldsymbol{v}}}{\omega_r - \boldsymbol{k} \cdot \boldsymbol{v} + i\gamma} \tag{14}$$

For the general case of a Maxwellian plasma distribution

$$f_j(v) = (m_j/2\pi T_j)^{3/2} exp(-m_j \boldsymbol{v}^2/2T_j) \tag{15}$$

the integration can be carried out and the highest frequency normal modes are determined to be thermally modified electron plasma oscillations with approximate dispersion given by the real frequency, $\omega_r^2 \simeq \omega_{pe}^2(1 + 3k^2\lambda_D^2)$, and the Debye length(squared) is $\lambda_D^2 = T_e/4\pi n_e e^2 = v_{Te}^2/2\omega_{pe}^2$. The Landau damping increment is given by

$$\gamma_L = -\Big(\frac{\pi}{8}\Big)^{1/2} \Big(\frac{m_e}{T_e}\Big)^{3/2} \frac{\omega_r^4}{k^3} exp\Big(-\frac{m_e}{2T_e}\frac{\omega_r^2}{k^2}\Big) \tag{16}$$

which reduces to

$$\gamma_L = -\Big(\frac{\pi}{8}\Big)^{1/2} \frac{1}{(k\lambda_D)^3} exp\Big(-\frac{1}{2k^2\lambda_D^2} - \frac{3}{2}\Big) \tag{17}$$

using the real frequency given earlier under the assumption $k\lambda_D < 1$. In order to obtain accurate damping rates the numerical solution to the exact dispersion relation, Eq.(14), is obtained for comparison to the simulations. This dispersion relation is given in Sect. 3.3 without the beam.

In order to simulate Landau damping as an initial value problem, using the δf-PIC model we begin by specifying the source term in the weight evolution equation, Eq.(8). From Eq.(15) this is given as

$$\frac{\frac{\partial f_o(v)}{\partial v}}{f_o(v)} = \frac{-2v}{v_T^2} \tag{18}$$

Next, the marker distribution, g(t=0), is initialized. It is arbitrary but a convenient choice is one which is consistent with the initial distribution,f(v). Therefore, g(t=0)=f(t=0). For the Landau damping simulation we use a one-dimensional system with the following parameters. Particle number, $N = 5 \times 10^4$, normalized thermal velocity, $\tilde{v} = v_{Te}/\omega_{pe}\Delta = 2.5$, system size, $L = 32\Delta$, and time step, $\omega_{pe}\Delta t = 0.05$. An initial perturbation of the weights, of the form $\delta f/f_o = 1 + \varepsilon cos(k_o x)$, was given with $k_o\lambda_D = 0.5$ and $\varepsilon = 0.05$.

The results for the time evolution of the field mode energy,$k\lambda_D = 0.5$, are shown in Fig. 1. The mode energy oscillates at approximately the plasma frequency with an overall decay consistent with the analytical damping rate within a few percent. As can be seen, the low noise technique correctly describes linear decay down to very low amplitudes. For the same number of particles, the

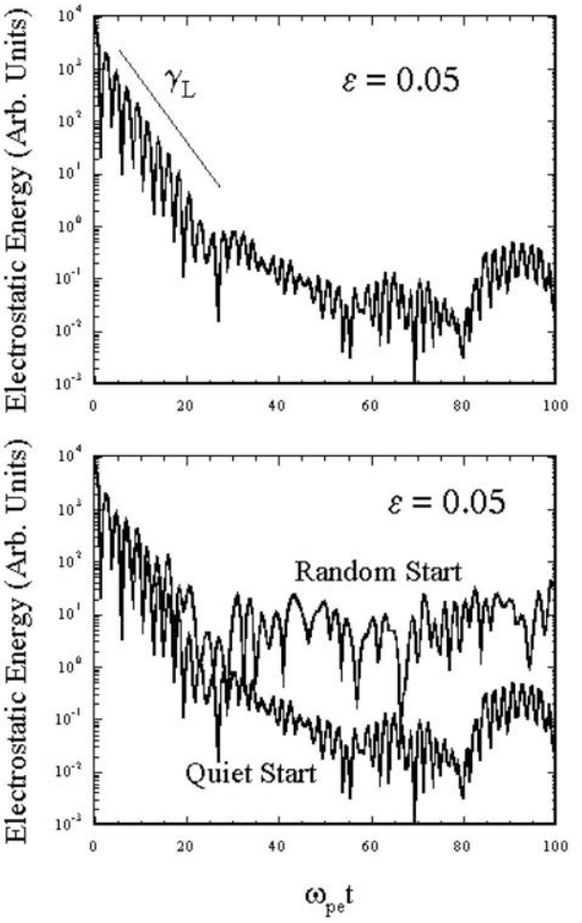


Fig. 1. Time evolution of the electric field mode energy for the case of linear Landau damping.

total-f PIC method shows Landau damping over about one order of magnitude and makes for a more difficult measurement. As the decay of the mode energy procedes, eventually there is a recurrence effect from free streaming where the density is restored to its initial value. The first recurrence peak occurs very close to the theoretically predicted value of $T_r \simeq 2\pi/k\Delta v \simeq 38\omega_{pe}^{-1}$.

The influence of marker distribution loading was also examined. In the first case we considered an ordered loading to cover the phase space. This is done using a bit-reversed quiet start method based on base-2 bit-reversed numbers [1,4]. This was compared to a random loading method consistent with the initial Maxwellian distribution. The comparison of the result for the mode energy time evolution is also shown in Fig. 1. The damping rates are consistent as

well as the first recurrence peak. The ordered marker start, however, showed a lower amplitude decrease and subsequent recurrence peaks were more clearly resolved. Therefore, a purely random marker distribution required more particles to achieve a comparable level of accuracy.

3.2 Nonlinear Landau Damping

Linear Landau damping involves resonant particles and the damping rate depends on the slope of the distribution function,f(v), at the phase velocity of the plasma wave. The distribution is considered unperturbed and this approximation is justified if the characteristic time of variation in f(v) is much larger than the wave damping time set by the Landau damping rate, γ_L^{-1}. For larger amplitude waves some of the particles traveling near the phase velocity will be trapped and oscillate in the quasi-stationary potential well. Nearly trapped particles will also exchange energy with the wave and as the wave amplitude changes, initially trapped particles may become untrapped. This makes for a rather complex evolution of the distribution function; eventually the initial wave damps to a finite steady amplitude with a zeroth order distribution function. This evolution has been studied previously by analytical methods and continuum Vlasov simulations [2,22,9]. It is considered a rigorous test of Vlasov simulations since the code uses a velocity grid and must accurately resolve detailed structural changes in the distribution function to a high degree of accuracy. The δf-PIC method quite easily describes this nonlinear damping accurately with relatively few marker particles.

The transition to the regime of nonlinear Landau damping can be estimated by considering the bounce period of the resonant electrons trapped by the wave electric field. This bounce period, given by $\tau_B = (2m_e/ek^2|\phi_k|)^{1/2}$ with ϕ_k the electric potential, is related to the characteristic time scale of the distribution function in the resonant velocity region. When $\gamma_L(k)\tau_B < 1$, the wave amplitudes are considered large and a threshold limit can be estimated as

$$|\phi_k| \simeq \frac{m_e}{ek^2}\gamma_L(k)^2 \tag{19}$$

The theory for the nonlinear evolution is too involved for presentation here but the general features of the solution are that for short times, $t < \tau_B$, the untrapped particles dominate and lead to $\gamma(t) \simeq \gamma_L$. The initial amplitude follows the linear decay and after about a bounce period the wave energy stops decaying as the trapped particles begin reconstructing the wave [19]. The reconstructed finite amplitude wave is lower that the original amplitude with a rough magnitude change given by $\phi_k(t) = \phi_k(t = 0)(1 - \gamma_L(k)\tau_B)$. The variation in the mode energy is dominated by the trapped particle oscillations which eventually lose coherence with the wave. It is important to note that this phase mixing effect of the trapped particles removes the original Landau damping increment.

This basic result is obtained in the δf-PIC simulations shown in Fig. 2 where we use the same parameters as the linear Landau damping study but increase the original perturbation strength to $\varepsilon = 0.2$. The initial energy damps more rapidly

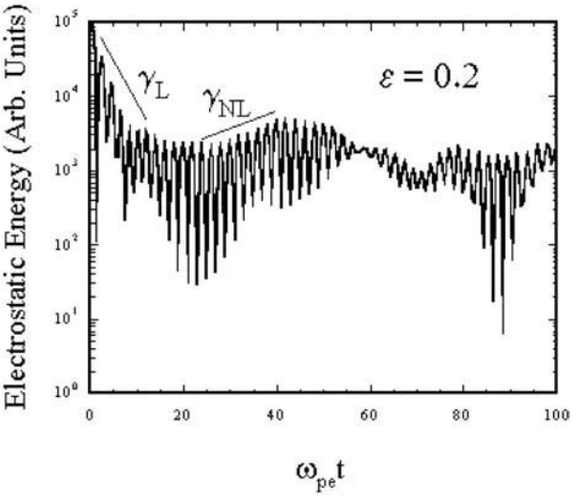


Fig. 2. Time evolution of the electric field mode energy for the case of nonlinear Landau damping.

than γ_L and reaches its first minimum at $\omega_{pe}t \simeq 18$. After this the mode energy grows exponentially until $\omega_{pe}t \simeq 40$ and then saturates. After saturation the basic mode oscillates with a period of about $20\omega_{pe}^{-1}$, which agrees well with the particle trapping time calculated from the electric potential at the first maximal value at $\omega_{pe}t \simeq 40$.

3.3 Weak Beam Instability

We now consider the case of a nonequilibrium distribution in the form of a background Maxwellian in the presence of a weak beam of particles with density, n_b. This is analytically described by

$$\begin{aligned} f_e(v) = (1-\epsilon)(m_e/2\pi T_e)^{3/2} exp(-m_e v^2/2T_e) + \\ \epsilon(m_e/2\pi T_b)^{3/2} exp(-m_e(v-v_b)^2/2T_b) \end{aligned} \tag{20}$$

where $\epsilon = n_b/n_e << 1$ and $|v_b| > v_{Te}$. We assume the ions are immobile and isotropic and one-dimensional motion for both beam and background electrons. In the absence of the beam, $\epsilon = 0$, electrostatic perturbations are damped according to the rate, γ_L, given previously. The presence of the beam gives rise to Landau resonances in this region of phase space and the possibility for wave growth due to the region of positive slope in f(v). The growth rate is given approximately by

$$\begin{aligned} \gamma = \gamma_L + \gamma_b \simeq -\Big(\frac{\pi}{8}\Big)^{1/2}\Big(\frac{m_e}{T_e}\Big)^{3/2}\frac{\omega_r^4}{k^3}(1-\epsilon)exp\Big(-\frac{m_e}{2T_e}\frac{\omega_r^2}{k^2}\Big) \\ + \epsilon\Big(\frac{T_e}{T_b}\Big)^{3/2}\Big(\frac{kv_b-\omega_r}{\omega_r}\Big)exp\Big(-\frac{m_e}{2T_b}\Big(\frac{\omega_r - kv_b}{k}\Big)^2\Big) \end{aligned} \tag{21}$$

where $\omega_r^2 \simeq \omega_{pe}^2(1+3k^2\lambda_D^2)$. The growth rate due to the beam is proportional to the beam density, ϵ, making this a challenging case for total-f PIC simulations in the weakly dense beam regime.

We now present results of the weak beam instability using both total-f and δf-PIC methods. The source term in the weight evolution equation is modified to become

$$\frac{\frac{\partial f_o(v)}{\partial v}}{f_o(v)} = \frac{-2v}{v_{Te}^2}(1-\epsilon) + \epsilon\Big(\frac{-2(v-v_b)}{v_b^2}\Big)\frac{f_b}{f_o} \tag{22}$$

The parameters used for this case were as follows; particle number, $N = 5\times10^5$, normalized thermal velocity, $\tilde{v} = 1.$, system length, $L = 32\Delta$, beam speed, $v_b = 5v_{Te}$, $T_e = T_b$, and beam density, $\epsilon = 10^{-5} - 10^{-3}$.

The linear growth rates quoted are only approximate and more precise values are obtained from the complete dispersion relation for the beam plasma system given by the following dielectric [14]

$$\epsilon_T(k,\omega_r+i\gamma) = 1 + \frac{\omega_{pe}^2}{k^2v_{Te}^2}\Big(1+\xi_e Z(\xi_e)\Big) + \frac{\omega_{pb}^2}{k^2v_{Tb}^2}\Big(1+\xi_b Z(\xi_b)\Big) \tag{23}$$

where $\xi_\alpha = (\omega - kv_{d\alpha})/kv_{T\alpha}$ and Z is the plasma dispersion function which is related to the complimentary error function with complex argument [14]. In Fig. 3 the growth rates versus wavenumber are plotted for different values of the beam density, ϵ. The maximum growth rate occurs at $k\lambda_D = 0.17$. Also shown is the maximum growth rate versus ϵ, decreasing nearly linearly.

Next, we present simulation results of the weak beam instability using the total-f PIC model. The initial distribution with weak beam is illustrated in Fig. 4a. Figure 4b shows the results of the total electrostatic energy versus time for beam densities of $\epsilon = 0.01$ and $\epsilon = 0.001$. 10^6 particles were used for these two cases and instability is observed for the larger beam density but not for the lower density. Over one order of magnitude more particles was needed to observe the weak growth in the low density case, however, the saturated state exhibited a high degree of noise. In Fig. 5 the results for the $\epsilon = 0.001$ case are shown using the δf-PIC model using 5×10^5 particles. Clear linear growth is observed over many decades and an accurate value of $\gamma = 0.015$ is obtained for the most unstable mode which is within 5 percent of the theoretical value. The post-saturation phase shows much clearer nonlinear trapped particle oscillations and frequency shifting.

3.4 Anisotropic Temperature-Driven Electromagnetic Instability

As a final example we consider an electromagnetic instability in the electron cyclotron range of frequency driven by thermal anisotropy [14,3]. A form of the plasma distribution which includes anisotropy is the bi-Maxwellian

$$f_{oe} = n_o\left(\frac{m_e}{2\pi T_\perp}\right)\left(\frac{m_e}{2\pi T_\parallel}\right)^{1/2} exp\left[-\frac{m_e v_\perp^2}{2T_\perp} - \frac{m_e v_\parallel^2}{2T_\parallel}\right] \tag{24}$$

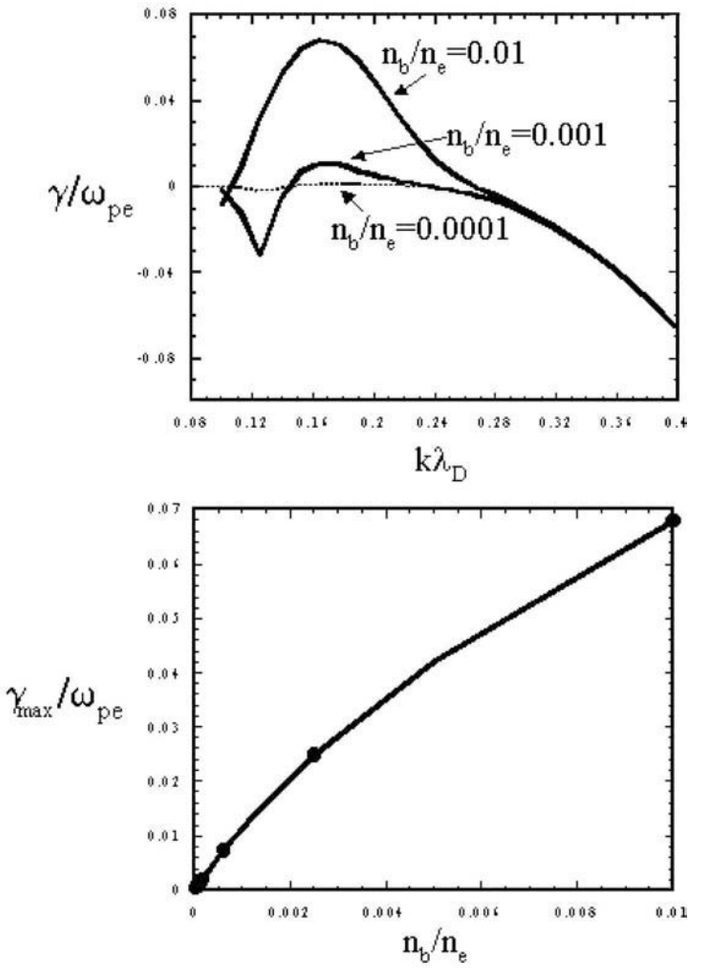


Fig. 3. Linear theory plots of the linear growth rate versus wavenumber for three beam densities and the lower plot is the maximum linear growth versus the beam density.

Parallel propagating, circularly polarized electromagnetic waves in the frequency range, $\omega_{ci} < \omega < \omega_{ce}$, where $\omega_c = qB/mc$ is the cyclotron frequency, are unstable when the electron distribution is characterized by $T_{\perp e} > T_{\|e}$. A similar instability exists for the ions but this case is not considered here. The linear dispersion relation for the parallel propagating modes is [14]

$$\epsilon_{\pm}(\omega,k) = 1 - \frac{k^2c^2}{\omega^2} - \frac{\omega_{pi}^2}{\omega^2} + \frac{\omega_{pe}^2}{\omega^2}\left[\frac{\omega_{ce}}{\sqrt{2}kv_{\|th}} Z\left(\frac{\omega \mp \omega_{ce}}{\sqrt{2}kv_{\|th}}\right) - \frac{T_\perp}{2T_\|} Z'\left(\frac{\omega \mp \omega_{ce}}{\sqrt{2}kv_{\|th}}\right)\right] \tag{25}$$

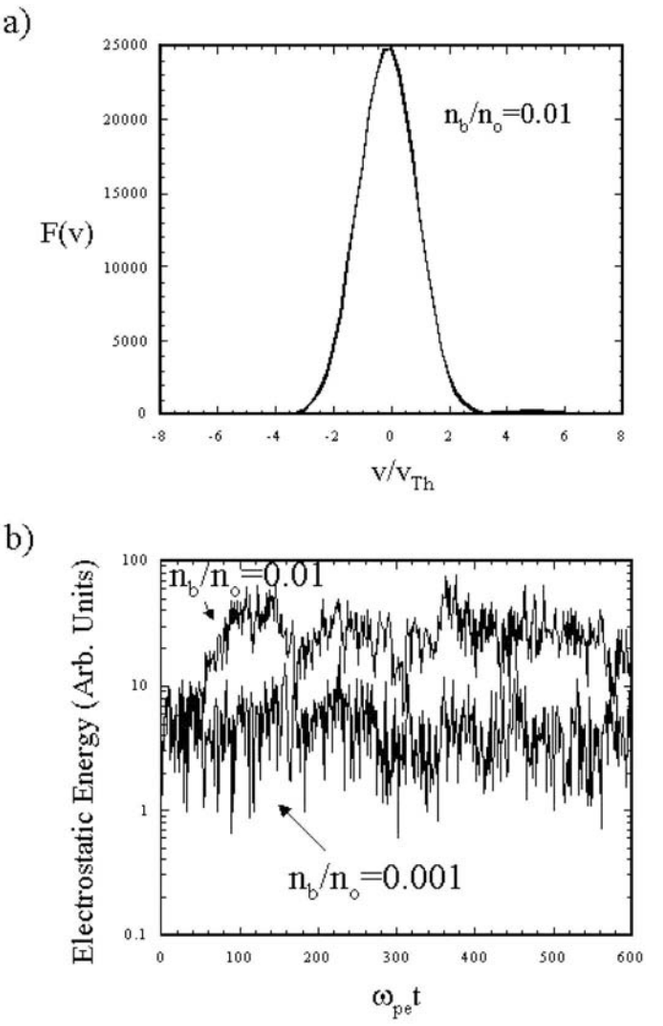


Fig. 4. Results from total-f PIC simulations. a) Initial electron distribution function for the weak beam instability case and the beam is barely visible at drift speed of $v_b = 5v_{Te}$ b) Total electric field energy for two different beam densities.

where $\pm$ refers to the right-handed(+) and left-handed(-) circularly polarized waves. The parameter k refers to the wavenumber parallel to the ambient magnetic field. For a moderately dense plasma, $\omega_{pe} > \omega_{ce}$, and perturbation frequency $|\omega_r|$ comparable to $|\omega_{ce}|$, the real frequency is approximated by $\omega_r \simeq$

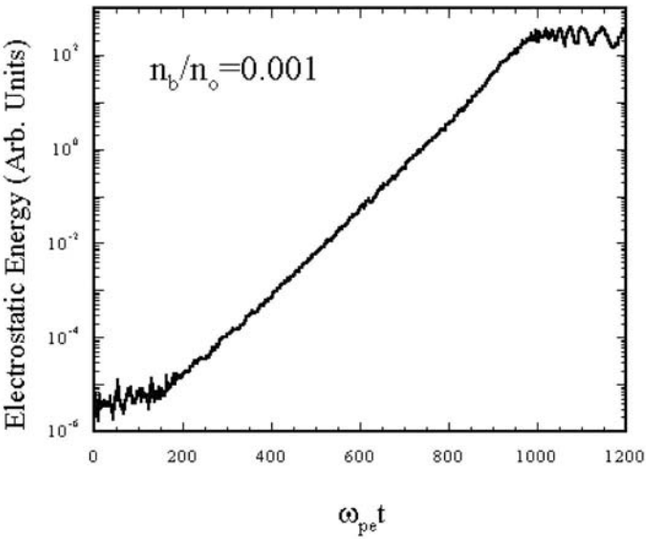


Fig. 5. Time evolution of the total electrostatic energy for the δf-PIC simulation of the weak beam instability.

$\mp\omega_{ce} \pm \omega_{ce}\omega_{pe}^2/(\omega_{pe}^2 + c^2k^2)$. The linear growth rate is given by

$$\gamma \simeq \sqrt{\pi}\frac{\omega_{pe}^2}{kv_{Te}}\Big(1 + \frac{c^2k^2}{\omega^2} + \frac{\omega_{pe}^2}{(\omega_r \pm \omega_{ce})^2}\Big)^{-1}\Big[-1 + \Big(1 - \frac{T_{e\perp}}{T_{e\parallel}}\Big)\frac{(\omega_r \pm \omega_{ce})}{\omega_r}\Big] exp\Big(-\frac{(\omega_r \pm \omega_{ce})^2}{k^2 v_{Te}^2}\Big) \tag{26}$$

For an isotropic plasma with $T_{\perp e} = T_{\parallel e}$, $\gamma < 0$, and the waves are weakly damped by resonant electrons with $kv_\parallel = \omega_r \pm \omega_{ce}$. For $T_{\perp e} > T_{\parallel e}$ the anisotropy causes wave growth and the threshold level for instability is given by the approximate condition

$$\frac{T_\perp}{T_\parallel} > \frac{1}{(1 - \omega_r/\omega_{ce})} \simeq 1 + \frac{c^2k^2}{\omega_{pe}^2} \tag{27}$$

For wavelengths $kc/\omega_{pe} \simeq 1$ the threshold anisotropy level is roughly 2. This marginal stability condition, which is only approximate, can be verified using the δf-PIC model since the noise level can be initialized to a very small level.

For the low noise electromagnetic simulations we use the system of equations given by Eq.(8). The source term for the weight evolution requires the gradient of the distribution which is

$$\begin{aligned}\frac{\nabla_{\boldsymbol{v}} f_o(v_\perp, v_\parallel)}{f_o} &= \frac{1}{f_o}(\hat{x}\nabla_{v_x} f_o + \hat{y}\nabla_{v_y} f_o + \hat{z}\nabla_{v_z} f_o) \\ &= (v_x\hat{x} + v_y\hat{y})\Big(\frac{\frac{1}{v_\perp}\partial_{v_\perp} f_o}{f_o}\Big) + \hat{z}\frac{\partial_{v_\parallel} f_o}{f_o}\end{aligned} \tag{28}$$

with $v_\perp = (v_x^2 + v_y^2)^{1/2}$ and $v_\parallel = v_z$. The derivative of the bi-Maxwellian is easily evaluated and substituted into Eq.(8). The electromagnetic simulations

are performed using one spatial and three velocity components. An external magnetic field is imposed in the direction of the spatial component and periodic boundary conditions are assumed. The simulation parameters used are listed as follows; system size, $L = 256\Delta$ with $\Delta/\lambda_D = 1$, collisionless skin depth, $c/\omega_{pe} = 16\Delta$, electron cyclotron frequency normalized to the electron plasma frequency, $\hat{\omega}_{ce} = 0.5$, $N_o = 10$ particles/grid cell, time step $\omega_{pe}\Delta t = 0.05$, and temperature anisotropy ranging from $T_{\perp e}/T_{\parallel e} = 1 - 9$.

The simulation results of the thermal anisotropy-driven electron cyclotron instability are shown in Fig. 6. The magnetic field modal energy versus time

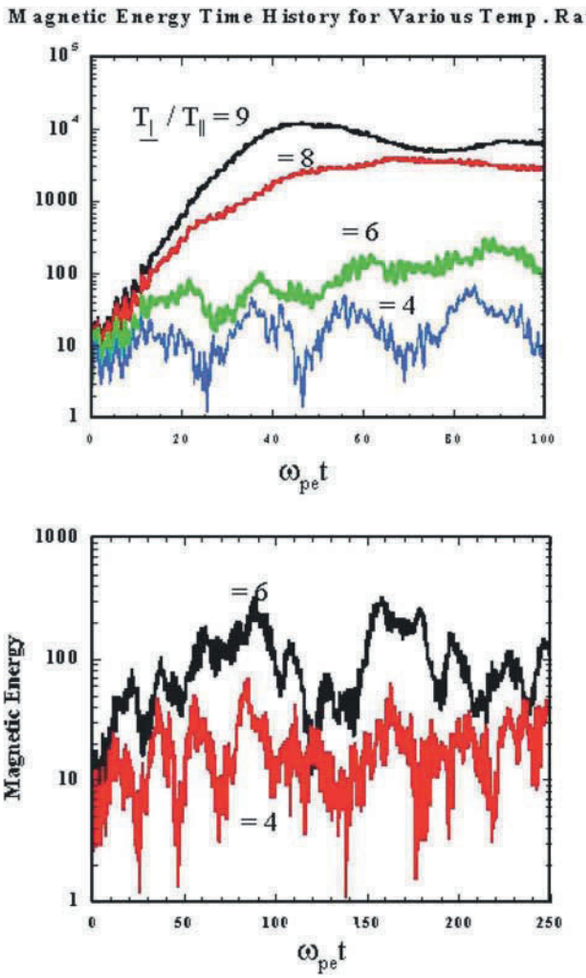


Fig. 6. Time evolution of the total magnetic energy for the electron temperature anisotropy-driven instability using the electromagnetic δf-PIC algorithm. The bottom panel is the extension of the simulations in the top panel for weaker anisotropy.

is plotted for various levels of initial anisotropy. The growth rate of the magnetic energy is reduced as the thermal anisotropy parameter approaches unity. The threshold level is determined to be $T_{\perp e}/T_{\|e} = 2.8$ which agrees well with the result obtained from the numerical solution to the complete dispersion relation($T_{\perp e}/T_{\|e} = 2.6$). Below an anisotropy level of 4 the total-f simulation could not resolve the linear growth rate accurately.

4 Summary

In this paper an efficient and accurate technique for study of the linear and nonlinear evolution of kinetic microinstabilities has been presented. This is achieved by following the evolution of the nonlinear perturbed plasma distribution, formulated in a Lagrangian approach which is suitable for particle simulation. The algorithm for the fully electromagnetic system was presented using an explicit time-centered form of the finite difference equations. For strongly unstable plasmas with $\gamma \geq \omega$ and where $\delta f/f \sim 1$, the conventional or total-f simulation approach is viable with large numbers of particles. In the opposite regime where instabilities are weak and saturation amplitudes low, the δf-PIC method works well. Several examples were used to illustrate this; linear and nonlinear Landau damping, the weak beam or bump-on-tail instability and the electromagnetic temperature anisotropy-driven electron cyclotron instability. In this latter case the marginal stability condition could be verified.

The total-f method is particularly difficult for investigating weak instabilities in multi-dimensional plasmas. Even if billions of particles can be used to represent the total distribution the integration time scales are very long to follow the growth and saturation dynamics. Therefore, only a limited regime can be investigated. Finally, the discrete nature of the δf-PIC method allows for straightforward parallelization of the algorithm, just as in the conventional PIC method. Hence, a large number of discrete markers can be used to cover the various regions of phase space and detailed particle convergence studies are possible.

Acknowledgements

This work is supported by a research grant from the Natural Sciences and Engineering Research Council(NSERC) of Canada.

References

1. C.K. Birdsall, A.B. Langdon: *Plasma Physics Via Computer Simulation* (McGraw-Hill, New York 1985)
2. C.Z. Cheng, G. Knorr: J. Comp. Phys. **22**, 330 (1976)
3. R.C. Davidson: *Methods in Nonlinear Plasma Theory* (Academic Press, New York 1972)
4. J. Denavit, J.M. Walsh: Comments Plasma Phys. Controlled Fusion **6**, 209 (1981)

5. R. Denton, M. Kotschenreuther: J. Comp. Phys. **119**, 283 (1995)
6. J.M. Dawson: Rev. Mod. Phys. **55**, 403 (1983)
7. A. Dimits, W.W. Lee: J. Comp. Phys. **107**, 309 (1993)
8. A.A. Galeev, R.Z. Sagdeev: *Nonlinear Plasma Theory* (W.A. Benjamin Press, New York 1969)
9. A. Ghizzo, P. Bertrand, M.M. Shoucri, T.W. Johnston, E. Fijalkow, M.R. Feix: J. Comp. Phys. **90**, 431 (1990)
10. A. Hasegawa: *Plasma Instabilities and Nonlinear Effects* (Springer-Verlag, New York 1975)
11. R.W. Hockney, J.W. Eastwood: *Computer Simulation Using Particles* (Adam Hilger Publishers, Bristol 1988)
12. G. Hu, J.A. Krommes: Phys. Plasmas **1**, 863 (1994)
13. M. Kotschenreuther: Bull. Amer. Phys. Soc. **34**,2107 (1988)
14. N.A. Krall, A.W. Trivelpiece: *Principles of Plasma Physics* (McGraw Hill, New York 1973)
15. A.B. Langdon, D.C. Barnes: 'Direct Implicit Plasma Simulation'. In: *Multiple Time Scales*, ed. by J.U. Brackbill, B.I. Cohen (Academic Press, Orlando, FL 1985) pp.335-375.
16. T. Miyake, Y. Omura, H. Matsumoto: J. Geophys. Res. **105**, 23239 (2000)
17. D.R. Nicholson: *Introduction to Plasma Theory* (John Wiley and Sons, 1983)
18. Y. Omura, H. Kojima, H. Matsumoto: Geo. Res. Lett. **21**, 2923 (1994)
19. T.M. O'Neil: Phys. Fluids **8**, 2255 (1965)
20. S.E. Parker, W.W. Lee: Phys. Fluids **B5**, 77 (1993)
21. P.L. Pritchett, 'Particle-in-Cell Simulation of Plasmas'. In: *Space Plasma Simulations*, this volume (Springer-Verlag, Heidelberg 2002)
22. M.M. Shoucri, R.R.J. Gagne: J. Comp. Phys. **27**, 315 (1978)
23. R.D. Sydora, V.K. Decyk, J.M. Dawson: Plasma Phys. and Contr. Fusion **38**, A281 (1996)
24. R.D. Sydora: Jour. Comp. and Appl. Math. **109**, 243 (1999)

Particle Simulation of Dusty Plasmas

Glenn Joyce, Martin Lampe, and Gurudas Ganguli

Plasma Physics Division Naval Research Laboratory
Washington DC 20375, USA

Abstract. The NRL Dynamically Shielded Dust (DSD) particle simulation code has been developed to model dusty plasmas. The code uses the techniques of molecular dynamics, particle-in-cell plasma simulations, and the test particle formulation of plasma kinetic theory. In addition, the interaction potential between dust particles is broken into long and short range parts using the P^3M technique of Hockney and Eastwood. The resulting simulation code avoids the necessity of integrating on plasma time scales, but still is able to include plasma effects.

1 Introduction

1.1 Experimental Properties of Dust

Dusty plasmas are ionized gases that also contain a distribution of charged solid particles, typically with particle diameters of 0.1 to 10μm. Under the most usual circumstances, the "dust" grains accumulate an electric charge which is negative and of the order of 10^3 to 10^4 times the charge of an electron, due to the preferential deposition of electrons on the grain surface. (It is also possible for the grain charge to be positive, for example in the presence of ultraviolet radiation which leads to electron emission from the particle.) Because of the very strong electrical interactions between the grains, and the mediation of these interactions by the plasma, dusty plasmas may exhibit new types of physical phenomena which in some sense represent a synthesis and extension of plasma physics, fluid physics and condensed matter physics.

Dusty plasmas occur in a vast variety of natural, experimental and commercial settings, including (among many others) the environment around space vehicles, the D-region of the ionosphere [1], and the rings of Saturn [2]. On the earth, they can also occur in discharges used for plasmas used in plasma processing [3]. Experiments on dusty plasmas generally proceed by deliberately introducing a "dust" of micron-sized particles into a discharge initiated between two electrodes and contained within a chamber [4]. At the electrodes and the chamber walls, there is always a sheath in which strong electric fields occur and serve to confine the plasma electrons within the chamber. These fields also drive a strong flow of the positively-charged ions toward the electrode. In an extended region called the presheath, the ion flow velocity builds up to the ion sound speed c_s, and within the sheath itself the flow speed exceeds c_s. Ion flows of this magnitude can strongly affect the properties of the plasma. In terrestrial experiments, the dust particles are supported against gravity by strong electric

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 125–135, 2003.

fields, and these fields strongly influence and limit the physical phenomena that can be studied. For example, the dusty plasma is usually confined to a thin sheet at the edge of the plasma sheath, and is subject to strong ion flows which are driven by the electric fields. The electric field at the sheath edge levitates the dust particles against gravity and confines them above the electrode. In this situation, the dusty plasma is found to be anisotropic and typically two-dimensional, because of the thinness of the dust cloud (frequently only a few layers of dust particles), and also because the ion flow profoundly influences the interactions between dust grains and establishes a preferential direction. In a microgravity environment, it is possible to fill large volumes with dusty plasma and to vary the electric fields over a wide range, and thereby to support dusty plasmas that have entirely different properties [5]. Dusty plasmas in microgravity may also ultimately have important commercial applications, for example in the deposition of thin films with unique properties. "New" plasma physics has emerged from experimental studies in terrestrial and microgravity environments with new modes of oscillation and other properties not seen previously in "normal" plasmas. If the dust is cooled enough, the dust particles may act as a liquid or in some cases as a solid [4]. One set of these new properties is concerned with the formation of Coulomb lattices, the lattice behavior, and, with the addition of kinetic energy to the dust, the melting of the lattice. Experimentally, the cooling occurs at neutral pressures above some critical pressure, P_{crit}. There is a wake downstream of each grain caused by the streaming ions. The wake effects cause positively charged regions behind the grains that can attract other grains. It is thought that these wakes are responsible for the lattice formation in the cold dust. When the neutral pressure is lowered below P_{crit}, the grains acquire a large random kinetic energy, and the dust "melts" forming a weakly coupled fluid.

1.2 Kinetic Modeling of Dust

The state of the modeling of dusty plasma phenomena is less developed than the state of experiments. Current kinetic models fall into two general classes: molecular dynamics models which follow the dynamics of the dust grains but normally employ a model to include the effects of the plasma, and particle-in-cell or fluid models which follow the plasma dynamics [6] [7] [8]. Because of the large difference in the time scales of the dust particles which may be on the order of 10^{-2} sec. and the plasma particles (the ion time scales are typically 10^{-7} sec.), particle-in-cell codes are time consuming, and generally are used to model non-linear plasma effects in the presence of stationary dust particles. There have been some studies including dust motion, but usually these are in fewer than three dimensions.

Molecular Dynamics. In "molecular dynamics" simulations, the force on any given particle is calculated by directly summing up the forces due to each of the other particles. This approach works well when interparticle forces are short-range so that each particle is influenced only by a small number of other particles in its immediate vicinity. In a plasma, however, long-range interactions arise

due to the slow fall-off of the Coulomb potential, and in a sense every charged particle is influenced by every other charged particle. It usually is not possible to do calculations following the actual number of electrons and ions (which may be a number N_{real} that lies somewhere between 10^{10} and 10^{22}), but often realistic simulations can be done using a number of particles N which is much smaller than N_{real}, but still very large (e.g., 10^6). However, two problems arise: the molecular dynamics approach requires N^2 calculations per time step which becomes prohibitive computationally, and the reduction from N_{real} to N simulation particles results in an unphysical enhancement of short-range "collisional" interactions.

One approach has been to use molecular dynamics simulations of the dust grains using the Debye-shielded Coulomb potential,

$$\varphi(\mathbf{k}) = \frac{qe^{-r/\lambda_{D_e}}}{r} \tag{1}$$

In this case, the plasma does not appear except as the source of Debye shielding. This force is, however, isotropic and does not lead to the types of strongly coupled structures seen experimentally.

Particle-in-Cell (PIC). In this approach, the charge of each particle is distributed among the nearest grid points, and then the electric field is calculated by solving Poisson's equation using the charge density on the grid as the source [9] [10]. In addition to greatly reducing the computational burden when N is large, this approach in effect smears out each point particle j, located at $\mathbf{r}_j$, into a finite-sized charge distribution with a specified shape $S(\mathbf{r}_j)$. This then eliminates the strong short-range interaction of charged point particles.

2 NRL Dynamically Shielded Dust (DSD) Simulation Code

In the study of dusty plasmas, neither of these approaches is completely satisfactory. In the strongly-coupled dust component, the short-range part of the interparticle interaction is of paramount importance in the formation of lattices, organization of vortices, and so on. However, the electrons and ions in the plasma are weakly coupled, and participate in long range collective interactions that play an essential role. The electron-ion plasma also serves as a medium which mediates long-range interactions between dust particles, and these are thought to also play an important role in dust structuring and dynamics. Furthermore, the dynamics of the dust occurs on a vastly slower time scale than that of the electrons and ions (because a dust particle is typically 10^{12} times the weight of an ion), and on a much coarser spatial scale (because the plasma particle density n_p may be three or more orders of magnitude higher than the dust density n_d). Because of these difficulties, complete simulations of the dusty plasma have not

been attempted; simulation studies have tended to focus on one aspect or another of the problem, and in many cases have made simple assumptions in order to elucidate qualitative aspects of the physics.

2.1 General Features

We have developed a code, DSD, to model strongly coupled dust in plasmas [13]. We make use of the techniques of molecular dynamics simulation, PIC simulation, the "particle-particle/particle-mesh" (P^3M) technique of Hockney and Eastwood [10] and we also make use of the dressed test particle representation, which is one of the theoretical foundations of plasma physics [11]. Many of the techniques we use in the model are common to all PIC plasma simulation codes. The unique properties of the code follow from the accurate representation of both the long-range and short-range aspects of the interaction between dust grains, as mediated by the complete plasma dielectric response.

In the usual case (near the sheath), the mean Coulomb potential energy of the electrons is $<<\mathrm{T}_e$, and the mean Coulomb potential energy of the ions is small compared to the ion streaming energy $\frac{1}{2}\mathrm{m}_i\mathrm{u}_i^2$. As a result, the onlinear perturbations to the ion orbits are weak and the response of the plasma to the presence of a test particle (in this case the "dust") is linear and is given by

$$\varphi_{true}(\mathbf{k}) = \frac{q}{2\pi^2 k^2 D(\mathbf{k}, \mathbf{k}\cdot\mathbf{v}_D)} \tag{2}$$

where $\mathbf{v}_D$ is the drift of the ions through the dust. Since the plasma response is linear, the potential due to many particles is the sum of the potentials of the individual particles. The plasma dielectric function, $\mathrm{D}(\mathbf{k},\omega)$ contains the plasma properties and for warm electrons can be written as

$$D(\mathbf{k},\omega) = 1 + \frac{1}{k^2\lambda_{De}} + \frac{\omega_{pi}^2}{k^2}\frac{\int d^3\mathbf{v}\frac{\mathbf{k}\cdot\partial f_{io}(\mathbf{v})/\partial\mathbf{v}}{\omega-\mathbf{k}\cdot\mathbf{v}+i\nu_i}}{1-i\nu_i\int d^3\mathbf{v}\frac{f_{io}(\mathbf{v})}{\omega-\mathbf{k}\cdot\mathbf{v}+i\nu_i}} \ . \tag{3}$$

The phase speeds of streaming instabilites which are sometimes present in the dust-plasma system are much faster than the ion or electron speeds. There are no resonances with the ions or electrons due these waves and the plasma response function given by equation (3) provides a good representation of the plasma.

Each dust grain may be thought of as an independent particle "dressed" by the plasma. For no drift, the potential becomes the Debye-Huckel potential

$$\varphi_{true}(\mathbf{k}) = \frac{qe^{-r(1+T_e/T_i)/\lambda_{D_e}}}{r} \tag{4}$$

and exhibits Debye shielding. When a drift is present, the resulting potential is anisotropic. Figures 1 and 2 demonstrate the wakes in a plasma caused by an ion stream from left to right. The test particle is at the position z = 0 and $textbfr$ = 0. In Fig. 1, the potential shows a wake behind the particle. The negative regions demonstrate attractive as well as repulsive regions. Even in the absence

of collisions with neutrals, the wake decays away due to Landau damping. As the collisions with neutrals increase, the wake damping increases. Figure 2 shows that the attractive regions behind the particle are also in the radial direction. The resulting forces tend to make particles line up in the streaming direction.

The DSD code represents the dust as simulation particles interacting through the dressed particle potential.

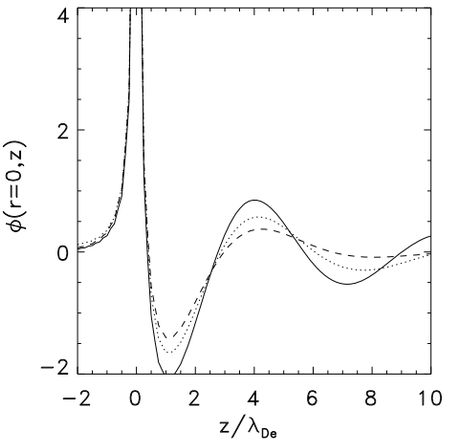


Fig. 1. Electrostatic potential of a stationary charge in a plasma with streaming ions. The ions are moving left to right. The solid line shows the potential with no collisions. The dashed and dotted lines show the effect of ion-neutral collisions on the potential. The collision frequency for the dashed line is higher than for the dotted line.

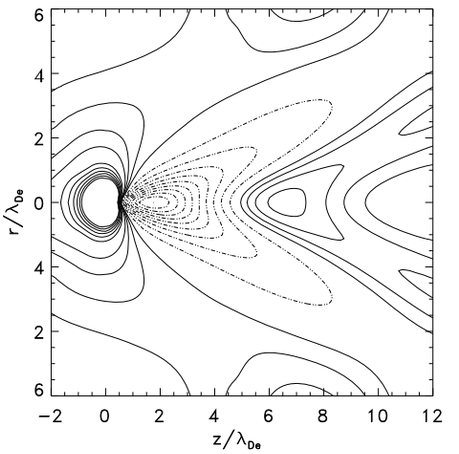


Fig. 2. Contour plot of the electrostatic potential of a stationary charge in a plasma with streaming ions. The solid contours denote regions of repulsion while the dashed contours represent regions of attraction.

2.2 Long and Short Range Interactions

Because of the laydown method, the particle-in-cell procedure gives correctly only long range force while molecular dynamics contains short-range interactions, but the dynamics requires the calculation of N^2 interactions. We use both techniques together with dressed particles according to the P^3M procedure of Hockney and Eastwood. The division of the particle potential into long- and short- range interactions using the P^3M allows a full force calculation while requiring a smaller number of interactions [10].

2.3 DSD Scheme in Detail

If $S(\mathbf{k})$ is the structure function of the finite-sized particles in a PIC model, i.e. the Fourier transform of the particle charge distribution $S(\mathbf{r})$, then in k-space the potential induced by the dressed PIC particle is

$$\varphi_{ref}(\mathbf{k}) = \frac{qS(\mathbf{k})}{2\pi^2 k^2 D(\mathbf{k}, \mathbf{k}\cdot\mathbf{v}_D)} \ . \tag{5}$$

This is the reference potential. Figure 3 shows the two particle interaction force using a linear laydown scheme in PIC. Note that when particles are closer than a cell length, the interparticle force decreases to zero linearly instead of increasing. The linear interpolation scheme has the property of treating the simulation particles as charge "clouds" instead of point particles. Particle-in-cell techniques do not treat the short-range forces correctly. To define the short range potential $\varphi_{sr}(\mathbf{k})$, we subtract the reference potential (the PIC potential) from the true potential; $\varphi_{sr}(\mathbf{k}) = \varphi_{true}(\mathbf{k}) - \varphi_{ref}(\mathbf{k})$ is zero outside a region near the particle. We may sum up directly the short-range force on any particle by all

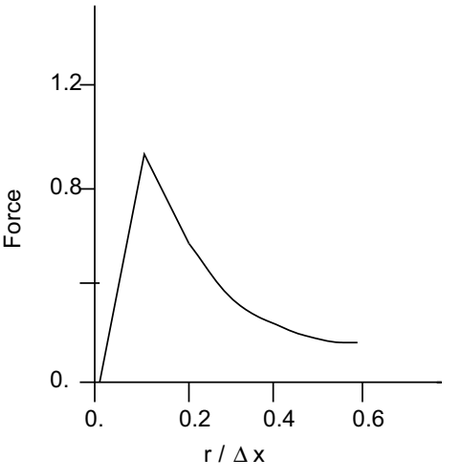


Fig. 3. Interaction force in two dimensions for linear laydown [12]

.

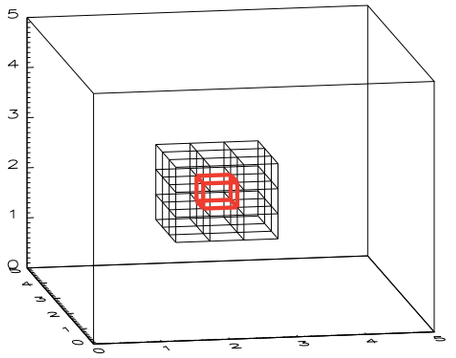


Fig. 4. Nearest neighbor cells in three dimensions.

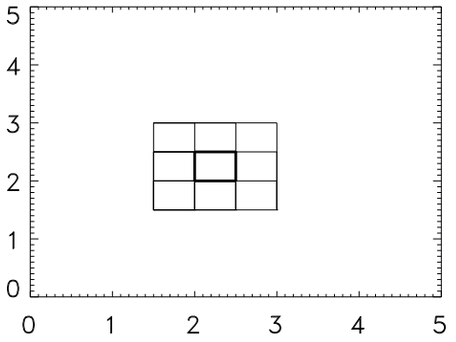


Fig. 5. Nearest neighbor cells in two dimensions.

other particles in its vicinity. Only a few particles in the nearest neighbor cells will contribute and this is not an N^2 operation. Figure 4 shows the region in which particles can contribute to the force on a particle in the interior cell. In order to be able to see this more clearly, Fig. 5 shows the nearest neighbors in two dimensions.

The method of pushing particle in the DSD code is to:

- Calculate the charge density on a mesh. For this, DSD uses linear weighting although higher order methods could be used.
- Calculate the reference force from the reference (PIC) potential, $\varphi_{ref}(\mathbf{k})$
- Calculate the momentum change from this force as is usually done in plasma simulations.
- Calculate the momentum change due to the short-range force using pair-wise interactions as is done in molecular dynamics simulations.
- Update the particle positions found from the resulting total momentum.

The advantage of this technique over treating the plasma ions as particles is that the time steps used in the code can be long compared to those required for simulating ions, while kinetic electron and ion properties such as wake effects due to streaming plasmas are included correctly in the force formalism. In addition to being efficient, the technique is very general, and can be made to include a first-principles treatment of all of the physical effects that are thought to be important. Scattering of ions and electrons off each other and off neutral atoms can be included in the dielectric function. Coulomb ion drag forces on the dust grains are included as a reaction force to the ion flow. Deposition of ion and electron charge on the dust grains can also be included, and a dipole component can be added to the charge distributions on a single dust grain. This model contains all of the physics necessary to simulate equilibrium states, phase transitions, excitations, and dynamics of the strongly-coupled dusty plasma in both ground-based and microgravity environments.

3 Examples

We now show a number of examples of the DSD code for representing strongly coupled plasmas in the laboratory.

The first example (Fig. 6) is a single layer of dust in a plane with no ion streaming. There is a confining force normal to the plane, and a linearly increasing force in the plane. This is an example of a two dimensional "Coulomb" crystal.

A second example (Fig. 7) is a single layer of dust in which the ions are streaming in the -z direction through the dust. In this example, the dust grains are aligned in the direction of the dust streaming.

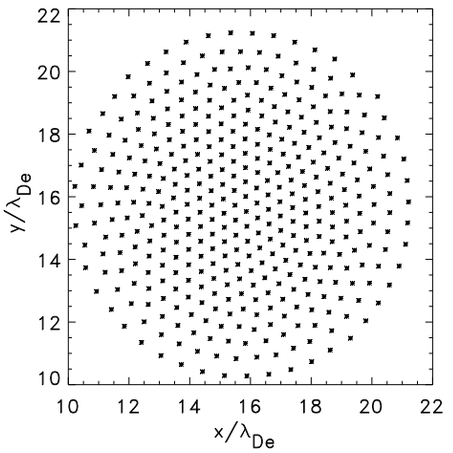


Fig. 6. Two dimensional Coulomb crystal simulated by the DSD code [13]

.

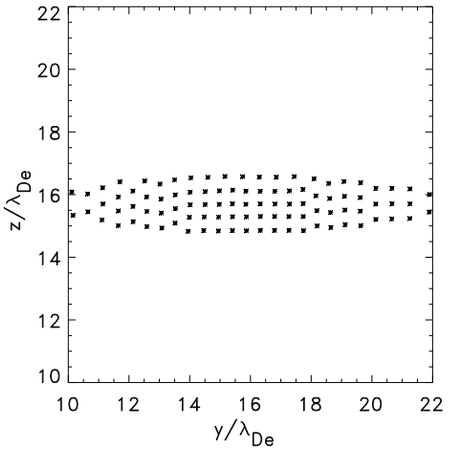


Fig. 7. Two dimensional dust layer in a plane. The plasma has streaming ions [13].

The last example (Fig. 8) is a three dimensional configuration of dust in a plasma in which the ions are streaming in the -z direction. In the upper plots (a) and (b), the neutral dust pressure is low, and the dust is hot. The left plot (a), shows that the particles circulate more or less freely within the dust cloud, and the right plot (b), shows no alignment between particles the upper part of the system and those in the lower part. In the lower plots (c) and (d), the neutral dust pressure is high enough that the dust particles are cold. The left plot (c), shows that the particles are confined to three planes with a few particles above and below. The right plot (d), shows that many of the particles in the top plane are aligned with those in the bottom plane. For more information on these phenomena see the article by Joyce, Lampe and Ganguli [14].

4 Summary

We have developed a simulation code, DSD, to model dust in a plasma. The code contains the elements of plasma kinetic theory, molecular dynamics, and particle-in-cell simulation. We have avoided having to resolve short time scales since the plasma appears only implicitly. The plasma is represented as a dielectric and contains the effects of ion streaming, Landau damping, and collisions of the ions with the neutral gas in the plasma and with each other. We have taken into account long and short range interactions using the P^3M method. The only particles in the computation are the dust particles. Using the DSD simulation model, we have been able to reproduce many of the phenomena associated with dust in a streaming plasma.

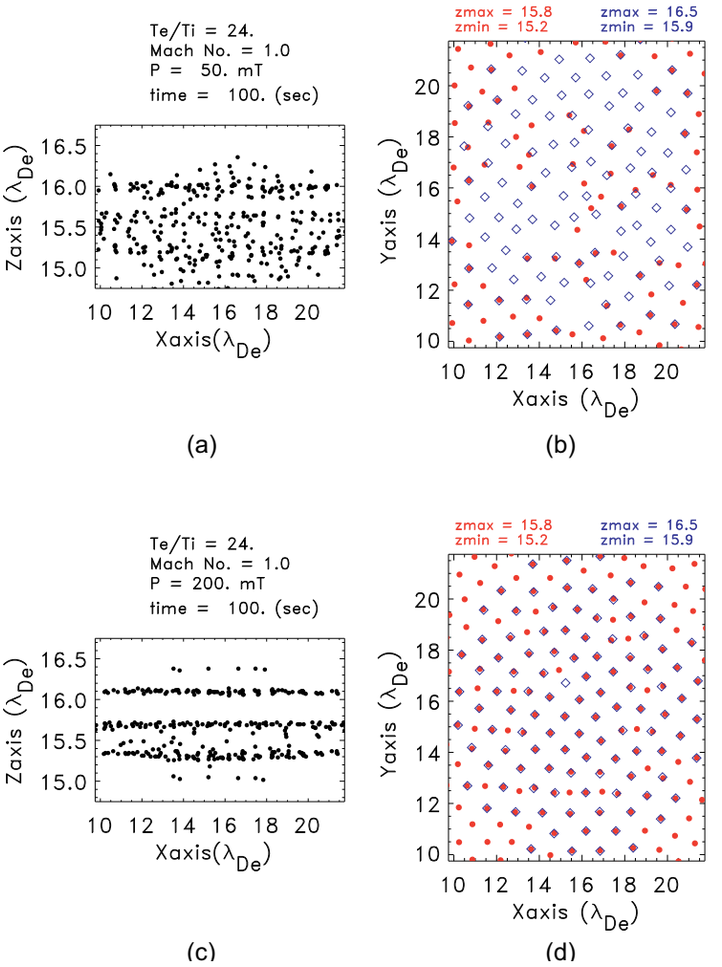


Fig. 8. Snapshot of grain locations. (a) and (b) are for a simulation with P=50 mTorr with hot particles. (a) Projection of the grains onto the x-z plane. (b) Projection of the grains onto the x-y plane. The streaming direction is z. (c) and (d) are similar plots for a simulation with P=200 mTorr. In this case the grains are cold. Blue diamonds indicate grains located in the top region z>15.9λ_D Red circles indicate grains in the lower region z<15.8λ_D.

References

1. O. Havnes, U. De Angelis, R. Bingham, C. K. Goertz, G. E. Morfill, and V. Tsytovich, *J. Atmos Terr. Phys.* **52** 637 (1990)
2. C. K. Goertz, *Rev. Geophys.* **44** 1197 (1989)
3. G. S. Selwyn, *Jpn. J. Appl. Phys.* **32** 3068 (1993)
4. H. Thomas, G. Morfill, V. Demmel, J. Goree, B. Feuerbacher, and D. Moehlmann, *Phys. Rev. Lett.* **72** 652 (1994)

5. G. E. Morfill, H. Thomas, U. Konopka, H. Rothermel, M. Zuzic, A. Ivlev, and J. Goree, *Phys. Rev. Lett.* **83** 1598 (1999)
6. F. Melandso and J. Goree, *J. Vac. Soc. Technol. A* **14** 511 (1996)
7. D. Winske, M. S. Murillo, and M. Rosenberg, *Physics of Dusty Plasmas AIP Conf. Proceedings 446* Woodbury, New York (1998)
8. J. E. Hammerberg, D. S. Lemons, M. S. Murillo, and D. Winske, *IEEE Trans. on Plasma Science* **29** 247 (2001)
9. C. K. Birdsall and A. B. Langdon, *Plasma Physics via Computer Simulation* Bristol, U.K. Adam Hilger (1991)
10. R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* Bristol, U.K. Adam Hilger (1988)
11. N. Rostoker *Nucl. Fusion **1** 101 (1960)*
12. J. W. Eastwood and R. W. Hockney, *J. Comput. Phys.* **16** 342 (1974)
13. G. Joyce, M. Lampe, and G. Ganguli, *IEEE Trans. Plas. Sci.* **29** 238 (2001)
14. G. Joyce, M. Lampe, and G. Ganguli, *Phys. Rev. Lett.* **88** 095006 (2002)

Hybrid Simulation Codes: Past, Present and Future – A Tutorial

Dan Winske[1], Lin Yin[1], Nick Omidi[2,3], Homa Karimabadi[2], and Kevin Quest[2]

[1] Applied Physics Division, Los Alamos National Laboratory, Los Alamos, NM 87545 USA *winske@lanl.gov, lyin@lanl.gov*
[2] Electrical and Computer Engineering Department, University of California San Diego, La Jolla, CA 92093 USA *nomidi1@san.rr.com*
[3] California Space Institute, University of California San Diego, La Jolla, CA 92093 USA

Abstract. Hybrid codes, in which the ions are treated kinetically and the electrons are assumed to be a massless fluid, have been widely used in space physics over the past two decades. These codes are used to model phenomena on ion inertial and gyro-radius scales, which fall between longer scales obtained by magnetohydrodynamic simulations and shorter scales attainable by full particle simulations. In this tutorial, the assumptions and equations of the hybrid model are described along with some most commonly used numerical implementations. Modifications to include finite electron mass are also briefly discussed. Examples of results of two-dimensional hybrid simulations are used to illustrate the method, to indicate some of the tradeoffs that need to be addressed in a realistic calculation, and to demonstrate the utility of the technique for problems of contemporary interest. Some speculation about the future direction of space physics research using hybrid codes is also provided.

1 Hybrid Codes: Past

Generally, the term "hybrid code" in plasma physics can refer to any simulation model in which one or more of the plasma species are treated as a single or multiple fluids, while the remaining species are treated kinetically as particles. The plasma can be coupled to the electromagnetic fields in a variety of ways: full Maxwell equations, low-frequency magnetostatic (Darwin) model, electrostatic only, etc. In this tutorial, we shall concentrate mainly on the most common type of hybrid code used in space plasmas: where all the ions are treated kinetically, the electrons are assumed to be an inertia-less and quasi-neutral fluid, and the electromagnetic fields are treated in the low-frequency approximation. Some comments on the extension to finite electron mass hybrid algorithms will also be made.

Because this tutorial is being presented in the context of the International School for Space Simulation (ISSS), we will mostly restrict the discussion of "past" uses of hybrid methods in space physics to the articles published in the previous schools. Those articles give appropriate and timely references to research that was carried out at that time with hybrid codes that were then available.

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 136–165, 2003.

In the same vein, in this tutorial the principal focus will be on numerical implementations of the hybrid model that are presently being used, and similarities and differences among the various algorithms. Numerical examples are discussed, in the context of some research questions currently being investigated with hybrid codes. This is the subject of Sect. 2: Hybrid codes – present.

Simulation codes with particle ions and fluid electrons appeared in the early 70's. Auer et al. [1] treated the ions as charged sheets and studied shocks and associated heating in relation to the Earth's bow shock. Forslund and Freidberg [2] and Mason [3] modeled shocks with unmagnetized ions treated by the particle-in-cell method. Chodura [4] wrote the first true "hybrid" code in the sense described above and applied it to laboratory pinch experiments. The hybrid model was subsequently used by Sgro and Nielson [5] and Hamasaki et al. [6] for related types of experiments. A 2-D version of the model was worked out by Hewett [7]. Leroy et al. [8] then adopted the Sgro and Nielson algorithm to model the structure of the bow shock. From there, the variations of algorithms and applications for space physics grew rapidly, as discussed later in this paper.

Hybrid codes have been well represented in the publications from the previous Simulation Schools. One-dimensional hybrid algorithms are discussed in some detail in the articles from the first [9] and second [10] schools. These articles emphasize applications to collisionless shocks and low-frequency waves in the ion foreshock. At the third school, Quest [11] discussed hybrid codes more generally, comparing and contrasting several multi-dimensional hybrid algorithms. As we will see, the field has not advanced much beyond what was described in that article. Winske and Omidi [12] presented a tutorial on hybrid codes at the fourth school that once again emphasized, for pedagogical reasons, one-dimensional codes. In this case, ion beam instabilities were used to illustrate the main features of hybrid simulation methods. Given that rather complete treatment, the emphasis in this tutorial will be on two-dimensional implementations and applications. These applications include ion beam instabilities, magnetic reconnection, and global hybrid simulations of the solar wind-magnetosphere interaction. Students and young researchers who are not familiar with hybrid techniques are urged to consult that ISSS-4 article as a reference point to the following discussion. The Winske and Omidi tutorial concluded with an outlook of how hybrid codes might evolve in future years. We will use that discussion as the basis for gauging progress over the past ten years as well as for making another prediction for the years ahead in Sect. 3: Hybrid codes – future.

Hybrid codes are also discussed in other articles in this volume, based on the tutorials given at ISSS-6. In particular, Pritchett [13] presents a very lucid tutorial on particle-in-cell methods generally and discusses the hybrid model in the context of other plasma kinetic models. Articles in the proceedings of ISSS-6 also consider a number of current applications of hybrid codes [14]–[18].

While the emphasis here is on work presented at ISSS, we also note three recent review articles that are also of interest. Winske and Omidi [19] discussed the general issue of the use of kinetic simulations in space plasmas and what type of simulation model (full particle, hybrid, etc.) is most appropriate under

various conditions. The strengths and weaknesses of kinetic simulation methods in general are discussed and some guidance on how to discern between "good" and "bad" simulations is provided. A more recent review of simulation methods and space physics applications is given by Pritchett [20]. Omura et al. [21] summarized relevant work on nonlinear kinetic processes in space plasmas in the recent URSI quadrennial report. It discusses results of observations, theory and simulations for a wide variety of phenomena and contains results from a number of hybrid simulations. On rereading this article, what strikes us most significantly is how simulations, whether they be full particle, hybrid, or MHD, have become so intermeshed with theory and observations in space physics. When ISSS was established, this was the principal goal. We have often failed to realize how well we have achieved this objective in the last twenty years.

Hybrid codes with finite electron mass are an interesting extension. This subject has been covered in depth in a recent monograph [22] that also considers inertia-less hybrid models. Given that extensive treatment, we will present only a brief discussion of this modification in the middle of Sect. 2. We have not included any discussion of hybrid codes that treat the electrons implicitly, which allows inertia and/or kinetic effects to be easily added in [23], or that deal with electrostatic phenomena [24]. If the scope of the readers' interests extends beyond the narrow confines of this tutorial, we urge them to consult these other references, as well as the magnetic (e.g., in [25] and [26]), and inertial fusion [27] literature.

2 Hybrid Codes: Present

2.1 Basic Assumptions and Equations

Hybrid codes arise from the need to model phenomena that occur on shorter time and distance scales than can be treated by magnetohydrodynamics and yet do not resolve processes that occur on electron scales (e.g., electron gyro-radius and electron Debye length scales, inverse electron gyrofrequency and electron plasma frequency time scales). The relevant scales are then the ion gyro-radius and ion inertial spatial scales, and inverse ion gyrofrequency time scale. In space, these length scales typically are on the order of 10's to 100's of km and times on the order of seconds; these ion scales are readily resolved by satellite instrumentation. To model phenomena on these scales with a hybrid code, as contrasted with a Hall-MHD code, implies assumptions about the descriptions of the plasma ions and electrons as well as the electromagnetic fields.

To be consistent with the hybrid model, the ions are treated kinetically, i.e., using standard particle-in-cell techniques [12][13][20][28]. Each simulation ion (charge q_i, mass m_i) is subject to the usual equations of motion:

$$m_i \frac{d\mathbf{v_p}}{dt} = q_i(\mathbf{E} + \frac{\mathbf{v_p} \times \mathbf{B}}{c}) \,, \tag{1}$$

$$\frac{d\mathbf{x_p}}{dt} = \mathbf{v_p} \,, \tag{2}$$

where $\mathbf{E}$ and $\mathbf{B}$, which have values given on a spatial grid, are the electric and magnetic fields interpolated to the particle location. The updated particle information is collected at the grid points to determine the ion number density (n_i), charge density ($q_i n_i$), flow velocity $\mathbf{V}_i$ and current $\mathbf{J}_i = q_i n_i \mathbf{V}_i$. (We will assume one ion species throughout. For multiple species, one accumulates the quantities for each species separately and then adds them together to determine the total ion charge density and current.)

In order to eliminate kinetic electron effects, the electrons are treated as an inertia-less fluid ($m_e = 0$). The electron momentum equation is thus:

$$n_e m_e \frac{d\mathbf{V_e}}{dt} = 0 = -en_e(\mathbf{E} + \frac{\mathbf{V}_e \times \mathbf{B}}{c}) - \nabla \cdot \underline{\mathbf{P}}_e \,, \tag{3}$$

where $\mathbf{V_e}$ is the electron fluid velocity and $\underline{\mathbf{P}}_e$ is the electron pressure tensor. Ignoring effects on the electron Debye length scale further implies that the plasma is quasi-neutral, so that the electron and ion charge densities are equal:

$$en_e = q_i n_i \,, \tag{4}$$

where the electron charge is $-e$ and n_e is the electron number density. In Eq. (3), $\underline{\mathbf{P}}_e$ is almost always taken as a scalar, $\underline{\mathbf{P}}_e = p_e \underline{\mathbf{1}}$. Typically, an isothermal or adiabatic relation between the pressure and temperature is assumed. For simplicity, we have also left off resistive coupling between the electrons and ions; this adds a term $en_e \underline{\eta} \cdot \mathbf{J}$ to the right-hand side of (3), where $\mathbf{J}$ is the total current. The resistivity $\underline{\eta}$ is usually taken as a scalar with a constant coefficient. For momentum conservation, it requires adding $-e\underline{\eta} \cdot \mathbf{J}$ to the acceleration term in the ion particle equation of motion (1) as well. We will return to the issues of the electron pressure tensor and the resistivity later.

Finally, the electromagnetic fields are treated in the low frequency approximation: Ampere's law,

$$\nabla \times \mathbf{B} = \frac{4\pi}{c}\mathbf{J} = \frac{4\pi}{c} q_i n_i (\mathbf{V}_i - \mathbf{V}_e) \,, \tag{5}$$

using (4), and Faraday's law

$$\frac{\partial \mathbf{B}}{\partial t} = -c(\nabla \times \mathbf{E}) \,. \tag{6}$$

As is done in two-fluid codes, Eq. (5) is used to eliminate $\mathbf{V}_e$ in (3) and Eq. (6) is used to advance the magnetic field in time. Because $m_e = 0$, Eq. (3) can be solved for the electric field directly, so that no time advance of $\mathbf{E}$ is needed. The other Maxwell's equations, e.g., Poisson's equation, $\nabla \cdot \mathbf{E} = 4\pi(q_i n_i - en_e)$, is satisfied by virtue of the quasi-neutral approximation (4) and boundary conditions; likewise $\nabla \cdot \mathbf{B} = 0$ is also satisfied.

Further discussion of the underlying assumptions of the hybrid model is given in papers describing the various hybrid algorithms, which we will consider in the next subsection.

2.2 Numerical Implementations

The numerical implementation of the hybrid model in a simulation code is relatively straightforward. However, as we shall see, one piece is not entirely there, which leads to various ways to circumvent this problem. Hybrid codes in space are usually (but not always, e.g., in [29]), done on a rectangular grid. Usually, a staggered pair of grids is used, with the electric field, plasma density, current, and pressure located on the vertices of one grid, and the magnetic field on the vertices of the other grid. This implies that curl **E** will be given properly on the B-grid, while curl **E** will be correct at the location of the electric field.

Given this grid, the ion dynamics are done as in any PIC code: the fields are interpolated to the particles' positions to give the correct acceleration and after the ions are moved, the density and current can be redeposited back to the grid [12][13][20][28]. Typically, linear weighting is used; higher order splines have also been employed. As usual, the particle positions and velocities are leapfrogged in time. In other words, at time step N (denoted by the superscript), the particle positions ${\mathbf{x}_p}^N$ and the fields $\mathbf{E}^N$ and $\mathbf{B}^N$ are known, while the particle velocities are known at the half-time step, $N-1/2$. Given $\mathbf{E}^N$ and $\mathbf{B}^N$, at the beginning of the time step, the velocities can be advanced to time level $N+1/2$ and the particle positions to time step $N+1$:

$$ {\mathbf{v}_p}^{N+1/2} = {\mathbf{v}_p}^{N-1/2} + \frac{q_i}{m_i}(\mathbf{E}^N + \frac{{\mathbf{v_p}}^N \times \mathbf{B}^N}{c})\Delta t\,, \tag{7} $$

$$ {\mathbf{x}_p}^{N+1} = {\mathbf{x}_p}^N + {\mathbf{v}_p}^{N+1/2}\Delta t\,. \tag{8} $$

In the process the currents are collected at $N+1/2$ and the density at $N+1$. Note that Eq. (7) is implicit in ${\mathbf{v}_p}^{N+1/2}$, as ${\mathbf{v}_p}^N$ can be written as $1/2({\mathbf{v}_p}^{N-1/2} + {\mathbf{v}_p}^{N+1/2})$ (see Appendix A in [12]).

Most hybrid codes solve the field equations explicitly in time, although implicit methods do exist [7][15][22]. In an explicit scheme, the straightforward way to advance the fields is to use Faraday's law (6) and $\mathbf{E}^N$ to advance the magnetic field to level $N+1/2$

$$ \mathbf{B}^{N+1/2} = \mathbf{B}^N - \frac{c\Delta t}{2}(\nabla \times \mathbf{E}^N)\,. \tag{9} $$

Recall that the electron momentum equation (3) can be solved for the electric field, which using (5), can be written as

$$ \begin{aligned} \mathbf{E}^{N+1/2} &= -\frac{{\mathbf{V_i}}^{N+1/2} \times \mathbf{B}^{N+1/2}}{c} - \frac{\nabla p_e^{N+1/2}}{q_i n_i^{N+1/2}} - \frac{\mathbf{B}^{N+1/2} \times (\nabla \times \mathbf{B}^{N+1/2})}{4\pi q_i n_i^{N+1/2}} \\ &= \mathbf{F}(\mathbf{B}^{N+1/2}, n_i^{N+1/2}, {\mathbf{V_i}}^{N+1/2})\,. \end{aligned} \tag{10} $$

Since we have the ion current and the magnetic field at time level $N+1/2$, and the density (either collected directly at $N+1/2$ or use the average between values at N and $N+1$), we have all the information needed to evaluate $\mathbf{E}^{N+1/2}$.

With the newly evaluated $\mathbf{E}^{N+1/2}$, again Faraday's law can be used to push $\mathbf{B}$ to time level $N+1$

$$\mathbf{B}^{N+1} = \mathbf{B}^{N+1/2} - \frac{c\Delta t}{2}(\nabla \times \mathbf{E}^{N+1/2})\,. \tag{11}$$

However, the advance of $\mathbf{E}$ from time level $N+1/2$ to $N+1$ is not so straightforward. An examination of (10) shows that $\mathbf{B}^{N+1}$ and n_i^{N+1} are known, but not ${\mathbf{V}_i}^{N+1}$. The problem of implementing a good algorithm for hybrid codes that reduces to how to best calculate $\mathbf{E}^{N+1}$. Much of the rest of this subsection is devoted to discussing various ways that this has been accomplished.

Historically the first method, which continues to be widely used, is a predictor-corrector technique [30]–[33]. The basic idea is to: (i) make a prediction of the fields (denoted by primes) at $N+1$; (ii) advance the particles in the predicted fields in order to compute the ion source terms at time level $N+3/2$; (iii) use that current (and charge density) to compute predicted fields at $N+3/2$; and (iv) use the average of the electric field at $N+1/2$ and the predicted field at $N+3/2$ to get $\mathbf{E}^{N+1}$. In equation form, this procedure becomes the following four steps [31][32]:
(i) the predicted fields

$${\mathbf{E}'}^{N+1} = -\mathbf{E}^{N} + 2\mathbf{E}^{N+1/2}\,, \tag{12}$$

$${\mathbf{B}'}^{N+1} = \mathbf{B}^{N+1/2} - \frac{c\Delta t}{2}(\nabla \times {\mathbf{E}'}^{N+1})\,; \tag{13}$$

(ii) advance the particles to obtain predicted source terms, $\mathbf{V}_{\mathbf{i}}'^{N+3/2}$, $n_i'^{N+3/2}$;
(iii) compute the predicted fields at $N+3/2$

$${\mathbf{B}'}^{N+3/2} = {\mathbf{B}'}^{N+1} - \frac{c\Delta t}{2}(\nabla \times {\mathbf{E}'}^{N+1})\,, \tag{14}$$

$${\mathbf{E}'}^{N+3/2} = \mathbf{F}({\mathbf{B}'}^{N+3/2}, n_i'^{N+3/2}, {\mathbf{V}'}^{N+3/2})\,; \tag{15}$$

(iv) Determine corrected fields at $N+1$

$$\mathbf{E}^{N+1} = \frac{1}{2}(\mathbf{E}^{N+1/2} + {\mathbf{E}'}^{N+3/2})\,, \tag{16}$$

$$\mathbf{B}^{N+1} = \mathbf{B}^{N+1/2} - \frac{c\Delta t}{2}(\nabla \times \mathbf{E}^{N+1})\,. \tag{17}$$

In principle, the process could be repeated to improve the accuracy, but in practice this is almost never done. This method is still often used because it gives very good energy conservation and is rather robust. As we shall see later, however, there can be significant amount of short wavelength whistler noise generated by the application of this technique, which will require additional

measures to remove. It is evident that this technique will be somewhat slow, since one has to move the particles twice each time step. Less evident at this point in the discussion is that the short wavelength noise problem may also require a somewhat smaller time step.

The second type of method to advance the electric field to time level $N+1$ is an extrapolation of the ion flow velocity (or equivalently the ion current density) from $N+1/2$ to $N+1$ [34][35]. Since the other quantities are known at $N+1$ already, with an extrapolated $\mathbf{V}_i^{N+1}$, $\mathbf{E}^{N+1}$can be directly evaluated, and the time-stepping process can proceed to the next cycle. Intuitively this method may not be as accurate, but it is better than one might expect, as we will show quantitatively later when we discuss some simple examples, and is often used for many problems.

The extrapolation of the ion velocity can be done in several ways. First, by merely using the saved values of $\mathbf{V}_i^{N-1/2}$ and $\mathbf{V}_i^{N+1/2}$, one has [35]:

$$\mathbf{V}_i^{N+1} = \frac{3}{2}\mathbf{V}_i^{N+1/2} - \frac{1}{2}\mathbf{V}_i^{N-1/2}\,. \tag{18}$$

One can also keep $\mathbf{V}_i^{N-3/2}$ to improve the accuracy, so that the process becomes a 4th order Bashford-Adams extrapolation, which is the way this method was first introduced [34]:

$$\mathbf{V}_i^{N+1} = 2\mathbf{V}_i^{N+1/2} - \frac{3}{2}\mathbf{V}_i^{N-1/2} + \frac{1}{2}\mathbf{V}_i^{N-3/2}\,. \tag{19}$$

Alternatively, one can follow the philosophy of implicit plasma methods and advance a moment equation to give a better estimate of the ion current [11]. This method requires accumulating the ion pressure tensor and may need some additional enhancement for multi-ion species problems, but it has shown promise [36]:

$$\begin{aligned}\mathbf{V}_i{}^{N+1} = \mathbf{V}_i{}^{N+1/2} &- \frac{\Delta t}{2}(\mathbf{V}_i\cdot\nabla\mathbf{V}_i)^{N+1/2} \\ &+\frac{\Delta t}{2m_i n_i^{N+1/2}}(-\nabla p_e{}^{N+1/2} - \nabla\cdot\underline{\mathbf{P}_i}^{N+1/2} + \frac{\mathbf{J}^{N+1/2}\times\mathbf{B}^{N+1/2}}{c})\,.\end{aligned} \tag{20}$$

However, this method requires the accumulation of the ion pressure tensor and the evaluation of an advective derivative, which would seem to negate main advantages of using a hybrid code that has particles to calculate the effects of $\underline{\mathbf{P}}_i$ and the advection already.

In the CAM-CL method, which has become popular in recent years, the ion current is calculated by doing an extra half time step push using a mixed level evaluation of the electric field [37]:

$$\mathbf{V'}_i{}^{N+1} = \mathbf{V}_i{}^{N+1/2} + \frac{q_i\Delta t}{m_i 2}(\mathbf{E}^* + \frac{\mathbf{V}_i{}^{N+1/2}\times\mathbf{B}^{N+1}}{c})\,, \tag{21}$$

$$\mathbf{E}^* = \mathbf{F}(\mathbf{B}^{N+1}, n_i^{N+1}, \mathbf{V}_i{}^{N+1/2})\,. \tag{22}$$

Note that here, as in all of these extrapolation methods, only one push of the ions each time step is required.

There are also several other hybrid algorithms, which require iterations. The Horowitz algorithm [38] solves the following (time-centered) set of field equations iteratively for $\mathbf{E}^{N+1}$ and $\mathbf{B}^{N+1}$, given the source terms at time level N+1/2:

$$\mathbf{B}^{N+1} = \mathbf{B}^N - \frac{c\Delta t}{2}(\mathbf{E}^N + \mathbf{E}^{N+1})\,, \tag{23}$$

$$\mathbf{E}^{N+1} = -\mathbf{E}^N + \frac{1}{2}\mathbf{F}(\frac{\mathbf{B}^N + \mathbf{B}^{N+1}}{2}, n_i^{N+1/2}, {\mathbf{V}_i}^{N+1/2})\,. \tag{24}$$

One starts the iteration by assuming $\mathbf{E}^{N+1} = \mathbf{E}^N$; typically 5-10 iterations are needed to give convergence.

Another iteration method involves a combination of the velocity extrapolation and predictor corrector [39]. After the velocity is estimated to calculate $\mathbf{E}^{N+1}$, the particles can be moved in the new fields to give a better estimate for ${\mathbf{V}_i}^{N+1}$ that can then be used to recompute $\mathbf{E}^{N+1}$; the system can be iterated until the $\mathbf{E}^{N+1}$ is in some sense converged.

While the emphasis here so far has been on the advance of the electric field, it should also be noted that there have improvements in advancing the magnetic field. Because of the generation of short wavelength whistlers, the time step restricting the advance of the magnetic field is smaller than that for the ions (which is a small fraction of the ion gyroperiod). Writing Faraday's law (6) in terms of the electric field (10), we have

$$\frac{\partial \mathbf{B}}{\partial t} = -c(\nabla \times \mathbf{E}) = \nabla \times \mathbf{F}(\mathbf{B}, n_i, \mathbf{V}_i)\,. \tag{25}$$

The quantities on the right-hand side are all evaluated at time level $N+1/2$, except for $\mathbf{B}$. One can then advance this equation from N to $N+1$ with $\mathbf{B}$ on the right-hand side changing, using a 4th-order Runge-Kutta scheme [39]; i.e., (suppressing the last two arguments of $\mathbf{F}$)

$$\mathbf{B}^{N+\theta} = \mathbf{B}^N + \frac{\Delta t'}{6}({\mathbf{K}_1}^N + 2{\mathbf{K}_2}^N + 2{\mathbf{K}_3}^N + {\mathbf{K}_4}^N)\,, \tag{26}$$

$${\mathbf{K}_1}^N = -\nabla \times \mathbf{F}(\mathbf{B}^N)\,, \tag{27}$$

$${\mathbf{K}_2}^N = -\nabla \times \mathbf{F}(\mathbf{B}^N + \frac{\Delta t'}{2}{\mathbf{K}_1}^N)\,, \tag{28}$$

$${\mathbf{K}_3}^N = -\nabla \times \mathbf{F}(\mathbf{B}^N + \frac{\Delta t'}{2}{\mathbf{K}_2}^N)\,, \tag{29}$$

$${\mathbf{K}_4}^N = -\nabla \times \mathbf{F}(\mathbf{B}^N + \Delta t'{\mathbf{K}_3}^N)\,, \tag{30}$$

using a subcycled time interval, $\Delta t' = \Delta t/\theta$.

Another approach that is used in the CAM-CL algorithm is to keep two copies of the magnetic field, one at the full time step and one at the half time

step and leapfrog these along in time, again subcycling if needed [37]. The two field solutions can be averaged after a given number of time steps.

Finally, it should be mentioned that if one ignores the electron pressure, the electric field can be eliminated entirely, resulting in a simpler algorithm [40]. This method is noteworthy in that a Richardson extrapolation (to $\Delta t \rightarrow 0$) is used in the particle velocity advance equation to dramatically improve the energy conservation. Two solutions, one advanced one time step (Δt) and the other advanced two half time steps ($\Delta t/2$) are compared and used to determine what the particle velocity at the next time step should actually be. But this is somewhat expensive, as two sets of particles are needed.

Before going on to numerical examples, we digress a bit to discuss hybrid codes with finite electron mass. The inclusion of electron inertia changes the whistler dispersion relation at short wavelengths, so that the real frequency no longer varies as the square of the wavenumber. If one is interested in resolving spatial scales that are a small fraction of the ion inertial length, this change of the wave phase velocity can reduce numerical problems at short wavelengths, so that one may wish to consider including $m_e \neq 0$ effects. There are several ways that this can be accomplished. The recently published monograph by Lipatov [22] is an excellent reference where the details can be found.

Hewett and Nielson [41] include electron inertia effects by a process that involves separating the electron current into its longitudinal (curl-free) and transverse (divergence-free) components. An equation for the total current can be obtained by summing the electron and ion momentum equations. Taking the divergence of this summed equation, and using the quasi-neutral condition expressed as $\nabla \cdot \mathbf{J} = 0$, one obtains an expression that can be solved for the longitudinal electric field, i.e., a generalized Poisson's equation for quasi-neutral plasmas. In similar fashion, the longitudinal part of the electron current, being curl-free, can be expressed as a scalar potential, $\mathbf{J}_{\rm el} = -\nabla V$. The divergence of this expression yields a Poisson equation, $\nabla^2 V = -\nabla \cdot \mathbf{J}_{\rm el} = -\nabla \cdot \mathbf{J}_l = \nabla \cdot \mathbf{J}_i$; the right-hand side is evaluated directed from the particle ions. Thus, knowing the total current, the ion current, and the longitudinal electron current, one easily obtains the electron transverse current ($\mathbf{J}_{\rm et}$). $\mathbf{J}_{\rm et}$ is advanced in time and used to calculate the vector potential $\mathbf{A}$ (or alternatively the time derivative of $\mathbf{J}_{\rm et}$ is advanced and used to directly calculate the transverse electric field $\mathbf{E}_t$). The method is straightforward, but a number of Poisson solves are needed to separate the transverse and longitudinal parts of the current and calculate the electric field.

In recent years, it has become more common to include electron inertia effects solving a set of generalized field equations [15]. One derives this set of equations from (3), but where the left-hand side is not set to zero, along with Eq. (6), by assuming:

$$\hat{\mathbf{B}} = \mathbf{B} - \delta_e^2 \nabla^2 \mathbf{B}\,, \qquad \delta_e = (c^2 m_e / 4\pi e^2)^{1/2} \tag{31}$$

and taking the curl of Eq. (3) to give

$$\frac{1}{c}\frac{\partial \hat{\mathbf{B}}}{\partial t} = -(\nabla \times \hat{\mathbf{E}})\,, \tag{32}$$

where

$$\hat{\mathbf{E}} = -\frac{\mathbf{V}_e \times \mathbf{B}}{c} - \frac{\nabla \cdot \underline{\underline{\mathbf{P}}}_e}{en_e} - \frac{m_e}{e}(\mathbf{V}_e \cdot \nabla)\mathbf{V}_e\,. \tag{33}$$

This set of equations is not exact; we have dropped terms proportional to $m_e \partial n_e/\partial t$ and $m_e \partial \mathbf{V}_i/\partial t$. It is argued that on short, electron spatial scales, the ions are nearly immobile and the neglected density and ion velocity variations are small [15][22][42][43]. The nice feature of Eqs. (31–33) is that one again has one equation for the time advance of the (generalized) magnetic field and a second equation for the (generalized) electric field that does not contain an explicit time derivative. These equations can be advanced in time using the same methods discussed previously.

Finite electron mass hybrid codes have been applied to a variety of problems, ranging from instabilities, e.g., the lower hybrid drift instability [41], collisionless shocks and solar wind-comet interactions [15][22], and magnetic reconnection [42][43]. The various calculations show the need to include finite electron mass to excite some instabilities, to generate strong whistler turbulence at shocks, and to show the role of short-scale physics in the reconnection process. It should be noted, however, that the fluid electron approximation eliminates electron Landau damping that tends to suppress whistler growth at short wavelengths. Thus, one may still want to use additional smoothing or some resistivity (but keeping the resistive length small compared to the cell size) to reduce the amplitude of the short wavelength fluctuations. And the generalized equations for solving for the electromagnetic fields with $m_e \neq 0$ should be compared with solutions of the full equations [41] for realistic test problems to check that the neglect of some of the time-dependent ion terms is indeed valid.

We emphasize that all of the methods discussed in this section work with various degrees of success that may also depend on the problem under consideration. It is best to keep several algorithms in your toolbox and try them all out when encountering a new application.

2.3 Examples

We next compare results from hybrid simulations based on two commonly used hybrid algorithms for a simple test problem. We show the potential tradeoff of speed versus accuracy. We go on to discuss the issue of short wavelength whistler noise and ways to suppress it. We also discuss a more complex test problem (reconnection in 2-D), comparing results from hybrid and Hall-MHD simulations. The possible use of hybrid techniques for more global problems raises enough issues that we will consider it in a separate section (2.4 Global Hybrid Calculations).

For the first comparison, we use two versions of the hybrid code we employ at Los Alamos. One version uses the standard predictor corrector scheme. The other uses velocity extrapolation along with a subcycled advance of the magnetic field using a 4th-order Runga-Kutta integration. We consider relatively small systems (128×128 cells, 50 particles per cell) with periodic boundary conditions to provide some illustrative comparisons. The test problem concerns the excitation of unstable, oblique Alfvén/ion cyclotron waves driven by cold, relatively slow ion beams that are found in the plasma sheet boundary layer and the solar wind [44]–[48]. These waves appear in the upstream region of slow shocks in the magnetotail and can be responsible for ion heating that occurs both upstream and in the shock transition. They also can provide ion heating in the solar wind. For this problem, results of two-dimensional simulations are shown in order to compare plasma and wave quantities and the effect of varying numerical parameters.

Specifically, we assume two equal density ions beams streaming relative to each other along the magnetic field (x-direction) with velocities of $\pm V_A$ (the Alfvénvelocity based on the total ion density). The beams have $\beta_{i\|} = \beta_{i\perp} = 0.025$ (with β based on the total ion density, and parallel and perpendicular are in reference to the magnetic field direction), and $T_e = T_i$, with the electrons treated adiabatically with $\gamma = 5/3$. The system length in both the x and y directions is $48c/\omega_i$ with ω_i the ion plasma frequency based on the total ion density. The calculations employ 128×128 cells, with 25 particles per cell for each ion species. The simulations use a resistivity (normalized to $4\pi/\omega_i$), usually $\eta = 10^{-6}$, corresponding to a resistive length about 1% of the cell size. In the velocity extrapolation runs, the time step is $\Omega_i \Delta t = 0.05$ (Ω_i is the ion gyrofrequency); the predictor-corrector runs require a smaller time step, $\Omega_i \Delta t = 0.02$, because of the persistence of short wavelength fluctuations. The predictor-corrector simulation smoothes the source term during both the predictor and corrector cycles; the velocity extrapolation simulation smoothes the source terms once or twice each time step. To obtain the smoothed value of a quantity at each grid point, we take one-quarter of the original value plus one eighth of the values of the four nearest neighbors and one-sixteenth of the values of the next four nearest neighbors.

Figure 1 shows results of two hybrid simulations comparing the velocity extrapolation (solid curves) and predictor/correct (dashed curves) algorithms. Plotted are time histories of the magnetic field fluctuations in the top panel (normalized to the ambient magnetic field) and the parallel and perpendicular ion temperatures (normalized to their initial values) in the middle panel. The magnetic fluctuations grow exponentially, saturate a modest level, $(\delta B/B_o)^2 \simeq 0.08$, oscillate, and then eventually decay. The ions are heated strongly in the perpendicular temperature, and much less in the parallel direction, consistent with earlier 1-D and 2-D simulations of this instability [45]. The power spectra of the fluctuations at $\Omega_i t = 100$ as a function of k_x are shown in the bottom panel. The spectra exhibit a broad peak of waves at long wavelength, with a steady fall-off at higher wavenumbers. The unstable waves grow slightly earlier in time in the predictor-corrector run, and show similarly time-shifted nonlinear oscillations

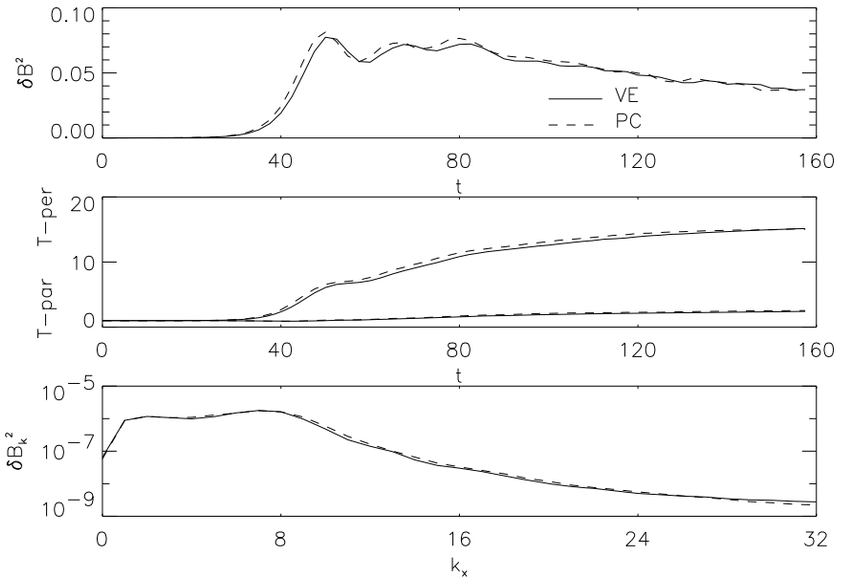


Fig. 1. Results of simulations of the Alfvénion beam instability comparing a velocity extrapolation algorithm (solid curves) with a predictor/corrector algorithm (dashed curves), showing magnetic field fluctuations (top panel), parallel and perpendicular ion temperatures (middle panel), both as a function of time, and power spectra of the magnetic fluctuations as a function of k_x at $\Omega_i t = 100$. Both calculations use the same normalized resistivity, $\eta = 10^{-6}$.

after saturation. The predictor-corrector run also shows slightly more perpendicular ion heating and a slight enhancement to the fluctuation level (except at the shortest wavelengths where it is reduced by the extra smoothing in this case).

Figure 2 compares profiles of the B_z magnetic field component (normalized to B_o) versus x at $y = L_y/2$ at various times for these simulations. At $\Omega_i t = 50$, at the end of the period of exponential wave growth, the wave profiles are regular in shape and very similar in the two calculations. By $\Omega_i t = 100$, the instability is into the nonlinear stage, but one can still see correspondence between the various peaks in the wave profiles, which have evolved into rather complex wave-forms. At $\Omega_i t = 150$, the wave spectra are evidently dominated by longer wavelength modes and the correspondence between individual peaks has been phase-mixed away to a large degree.

Figure 3 compares results for various velocity extrapolation runs in which the resistivity and/or the smoothing is changed. The solid curves correspond to the same case as shown in Fig. 1. In this run, the normalized resistivity is $\eta = 10^{-6}$ and the source terms are smoothed once each time step. A second run in which a second smoothing operation is included each time step is shown as the dashed lines. In this case, the excitation and growth of the fluctuations

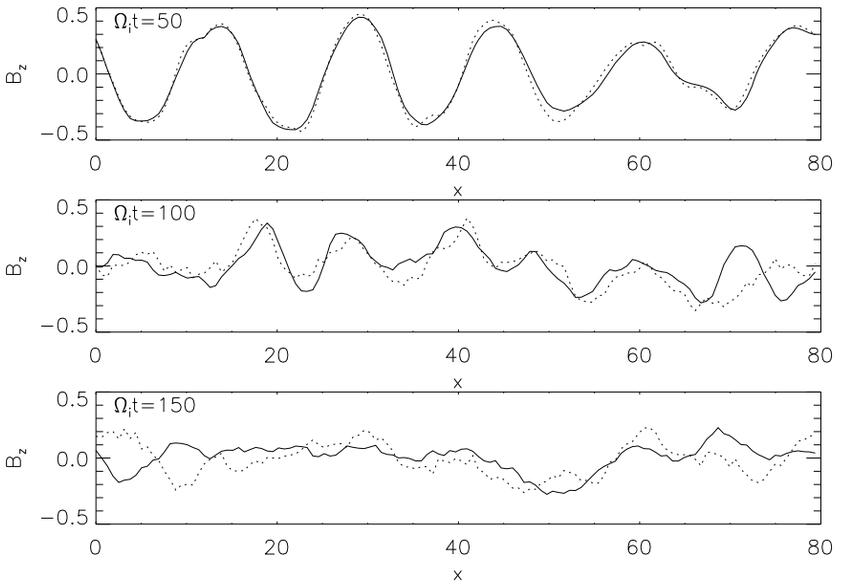


Fig. 2. Further results of the simulations given in Fig. 1, showing profiles of B_z versus x at $y = L_y/2$ at various times; solid curves correspond to the run using the velocity extrapolation algorithm, dotted curves to the predictor-corrector run.

is somewhat delayed due to the extra smoothing. In the nonlinear regime, the extra smoothing gives rise to enhanced magnetic field fluctuations at $\Omega_i t = 100$, which occur at longer wavelengths (as shown in the bottom panel). The third set of dotted curves correspond to a run in which two smoothes are employed and the resistivity is reduced by a factor of 10. In this case, there is a larger level of short wavelength fluctuations and consequently more perpendicular ion heating. At late times, the short wavelength modes increase the overall fluctuation level dramatically, as shown in the top panel. (With only one smoothing operation at this value of the resistivity, the short wavelength modes grow up earlier and the calculation does not run to completion.) Energy conservation was monitored for the runs. In the first case (solid curves), total energy (minus the energy in the uniform magnetic field) decreases by 1.3% by the end of the run. In the second case, with extra smoothing, the energy decrease is reduced to -0.75%. In the third case, in which the resistivity is reduced, the total energy increases by 9.6%, evidently due to the enhanced levels of short wavelength fluctuations and associated ion heating.

Figure 4 shows similar tests for predictor-corrector runs in which only the resistivity is changed. The dashed curves correspond to the same case shown in the Fig. 1 with $\eta = 10^{-6}$; the dashed lines correspond to a run in which the resistivity is reduced by a factor of 10. The time histories of the magnetic field fluctuations and parallel ion heating agree quite well in the two calculations. The perpendicular ion heating increases by about 20% when the resistivity is

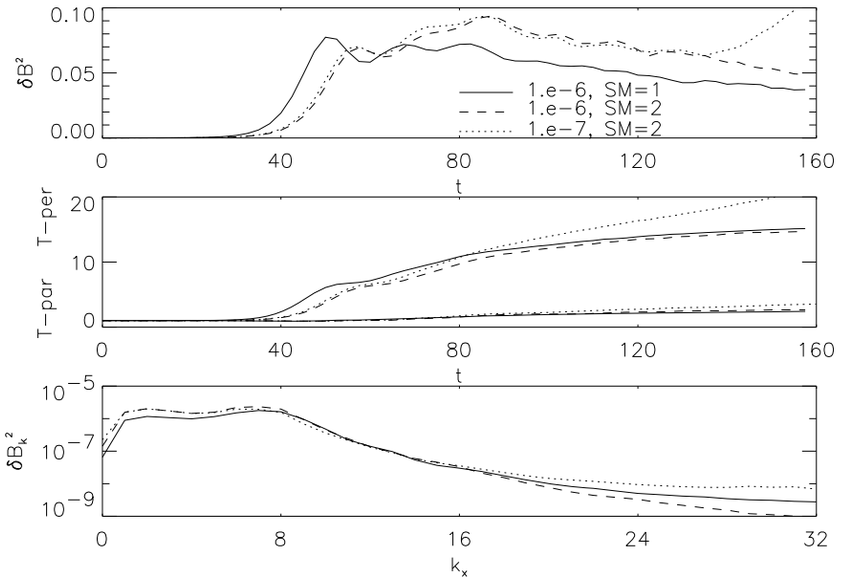


Fig. 3. Comparison of simulation results using the velocity extrapolation algorithm, in the same format as Fig. 1 for three runs: (solid curves) $\eta = 10^{-6}$ with one smoothing of the source terms each time step, (dashed curves) $\eta = 10^{-6}$ with two smoothes, and (dotted lines) $\eta = 10^{-7}$ with two smoothes.

reduced, again due to the increased level of short wavelength fluctuations, as shown in the bottom panel. Energy in the system for the higher resistivity case is reduced by about 2.9% by the end of the run. In the lower resistivity case, due to the excitation of short wavelength modes, the total energy increases by about 4.5%.

For a second test problem, we compare results of 2-D hybrid simulations with those from Hall-MHD calculations for reconnection in a (Harris) current sheet. As one can see from the generalized Ohm's law (3), the electric field responsible for reconnection can arise from any of the last three terms:

$$\mathbf{E} = -\frac{\mathbf{V_e} \times \mathbf{B}}{c} - \frac{\nabla \cdot \underline{\mathbf{P}}_e}{q_i n_i} + \eta \mathbf{J} - \frac{m_e}{e} \frac{d\mathbf{V_e}}{dt} \,. \tag{34}$$

In many hybrid calculations with $m_e = 0$, one uses a localized resistivity so that the consequences of reconnection in 3-D [49] or in a large 2-D tail configuration [50]–[54] can be explored. We have already discussed how to include the last term if $m_e \neq 0$. If one chooses to assume $m_e = 0$ and not to insert a large localized resistivity, one needs to include, in both the hybrid (and Hall-MHD) calculations, the full electron pressure tensor in (34) in order to initiate the reconnection process [55]–[57]. The electron pressure is advanced via the following

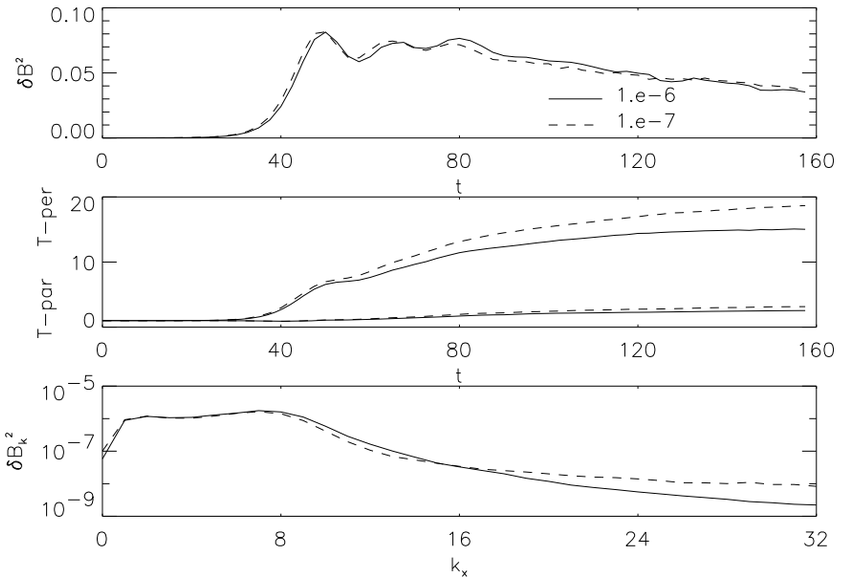


Fig. 4. Comparison of predictor/corrector calculations, in the same format as Fig. 1, for runs with normalized resistivities of $\eta = 10^{-6}$ (solid curves) and 10^{-7} (dashed curves).

equation that comes from a moment expansion of the Vlasov equation:

$$\frac{\partial \underline{\mathbf{P}}_e}{\partial t} = -\mathbf{V_e} \cdot \nabla \underline{\mathbf{P}}_e - \underline{\mathbf{P}}_e \nabla \cdot \mathbf{V_e} - \underline{\mathbf{P}}_e \cdot \nabla \mathbf{V_e} - (\underline{\mathbf{P}}_e \cdot \nabla \mathbf{V_e})^T$$
$$-\Omega_e [\underline{\mathbf{P}}_e \times \hat{\mathbf{b}} + (\underline{\mathbf{P}}_e \times \hat{\mathbf{b}})^T] - \nabla \cdot \underline{\overline{\mathbf{Q}}}\,, \tag{35}$$

where $\hat{\mathbf{b}} = \mathbf{B}/B_o$ and the superscript T indicates the transpose matrix. The term in square brackets involves the electron mass; it can either be treated implicitly, or combined with the last term, which is then replaced by a phenomenological relaxation term that models rapid electron-scale processes that reduce the non-gyrotropicity:

$$\Rightarrow -\frac{\Omega_e}{\tau}(\underline{\mathbf{P}}_e - p_e \underline{\mathbf{1}})\,, \tag{36}$$

where $p_e = \frac{1}{3}\mathrm{Tr}(\underline{\mathbf{P}}_e)$ and $\tau \sim 1$ is on the order of electron cyclotron time.

The simulations are performed in the (x, y) plane in which $\hat{x}$ is the direction normal to the sheet and $\hat{y}$ is along the sheet. An initial perturbation to the Harris equilibrium $B_y = B_o \tanh\{[x - l_x/2]/\alpha\}$ is employ, where l_x is the width of the simulation domain and α is the half-thickness of the current sheet. The perturbed normal magnetic field $B_x \sim \sin\{2\pi[y - l_y/2]/l_y\} \exp\{-[x - l_x/2]^2/\alpha^2\}$, where l_y is the length of the simulation domain, ensures the system evolves into a nonlinear stage quickly. The boundary conditions are periodic on the y-boundaries;

on the x-boundaries the velocity and field components transverse to the local magnetic field (i.e., the x and z components) vanish.

For the hybrid simulations the spatial sizes are 10×10 $(c/\omega_i)^2$ (for a current sheet with half thickness $\alpha = 0.4\, c/\omega_i$) and 20×20 $(c/\omega_i)^2$ (for $\alpha = 0.8\, c/\omega_i$), both consisting of 128×128 grids; the initial perturbation amplitudes are $B_x/B_o = 0.15$ and 0.3, respectively. The lobe plasma beta is 0.2 and the initial ion-to-electron temperature ratio is $T_i/T_e = 5$. The simulations include a uniform background ion population (20 background ions per grid) whose temperature is the same as the sheet ions (an average of 20 sheet ions per grid). ω_i is the ion plasma frequency based on the asymptotic plasma density at the lobe.

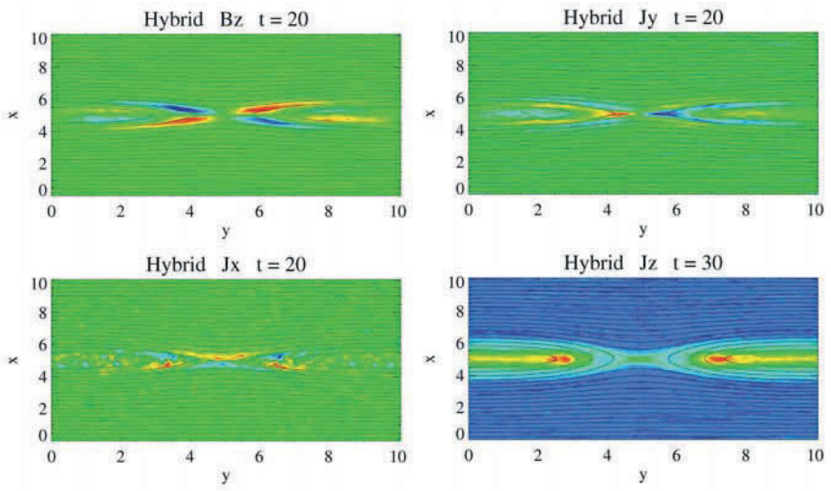


Fig. 5. Results of hybrid simulations of reconnection at $\Omega_i t = 20$ and 30 showing color contours of B_z and the three components of the total current; magnetic field lines are overlaid on the plot.

Figure 5 shows the magnetic field and current configuration during the nonlinear stage (at $t\Omega_i = 20$ and 30) from the hybrid run (the spatial size is 10×10 $(c/\omega_i)^2$ and $\alpha = 0.4\, c/\omega_i$). Overlaid on the magnetic field lines in the simulation plane are color contours of the out-of-plane magnetic field B_z and the current densities J_x, J_y, and J_z plotted on a linear scale.

For the B_z contour plots the direction of B_z is along $\hat{z}$ in the red regions but along $-\hat{z}$ in blue regions. The same color legend applies to the current density plots. Intense currents flow toward the X point along $\hat{y}$ and $-\hat{y}$ directions in localized regions at the center of the sheet (see the J_y plot) and diverge from the X point in the cross-sheet directions $\hat{x}$ and $-\hat{x}$ (seen near the X point in the J_x plot). The current pattern for J_x and J_y is consistent with the quadrupole configuration of the out-of-plane magnetic field B_z. As the reconnection proceeds, the out-of-plane current distribution evolves to a configuration in which

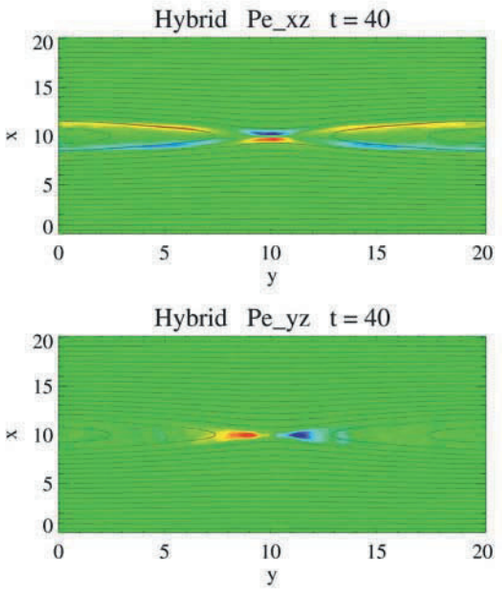


Fig. 6. Color contours of the xz and yz components of the electron pressure tensor from a hybrid calculation at $\Omega_i t = 40$.

J_z intensifies at localized regions in the center of the sheet away from the X point.

Global distributions of the electron pressure tensor terms $P^{\,e}_{\rm xz}$ and $P^{\,e}_{\rm yz}$ from the hybrid simulation (the spatial size is $20{\times}20\ (c/\omega_i)^2$, $\alpha = 0.8\, c/\omega_i$ at $\Omega_i t = 40$) are displayed in Fig. 6. The red and blue regions indicate maximum and minimum values, respectively. The magnitudes of the off-diagonal terms are relatively small: the peak values of these off-diagonal terms are a few percent of the peak values of the diagonal terms; yet their spatial configurations are clear. We will discuss their contribution to the reconnection electric field following the examination of the electric field configuration.

Next, we compare the hybrid and Hall-MHD simulation results for the 2-D reconnection problem. The Hall-MHD calculations use the same basic assumptions/equations and the same electric field and the electron pressure tensor model as in the hybrid calculations. Thus Eqs. (3) - (6) and Eqs. (32) and (33) remain the same. However, the ion kinetic treatment for each individual ion particle in Eqs. (35) and (36) is replaced with a fluid moment description in the Hall-MHD calculations. The momentum equation,

$$nm_i \frac{\partial \mathbf{V_i}}{\partial t} = -nm_i(\mathbf{V_i}\cdot\nabla)\mathbf{V_i} - \nabla p_i - \nabla\cdot\underline{\mathbf{P}}_e + \frac{\mathbf{J}\times\mathbf{B}}{c}\,, \tag{37}$$

describes the ion fluid motion. As in most fluid simulations, a scalar ion pressure p_i is used, and thus ion finite Larmor radius effects are not modeled. The plasma

density n is computed from the continuity equation

$$\frac{\partial n}{\partial t} = -\nabla \cdot (n\mathbf{V_i}) \,, \tag{38}$$

and p_i is obtained from

$$\frac{\partial p_i}{\partial t} = -\nabla \cdot (\mathbf{V_i} p_i) - (\gamma - 1) p_i (\nabla \cdot \mathbf{V_i}) \tag{39}$$

where $\gamma = 5/3$ is the ratio of the specific heats.

The Hall-MHD code used here was developed with numerical methods similar to those used in conventional hybrid codes [12][36][56] so that direct comparison of the results of the two codes could more easily be carried out. Spatial grid quantities in the code are defined on the same set of staggered finite-difference meshes as described in section B and the time differencing follows a conventional staggered leapfrog method. The magnetic field $\mathbf{B}$ is advanced in time using a fourth-order Runge-Kutta scheme with subcycling [36]. In each time step used to advance the fluid moments (velocity, density, and pressure), five to ten substeps are used, typically, to update the fields. A smoothing routine is used instead of the conventional second-order viscosity (i.e., $\sim \nu\nabla^2\mathbf{V}$); a sixth order hyperviscous dissipation, $\sim \nu(\nabla^6{}_x + \nabla^6{}_y)\mathbf{B}$, is employed in the magnetic field equation to damp out fluctuations on very short spatial scales. The Hall-MHD simulations use the same boundary conditions and simulation parameters. Since the fluid code runs relatively quickly, 256×256 grids are used for the 20×20 $(c/\omega_i)^2$ run for a better resolution.

Figure 7 shows a comparison of the global configuration of the electric field components E_y and E_z obtained from the hybrid run (left panels) and the Hall-MHD run (right panels). Since reconnection proceeds faster in the Hall-MHD

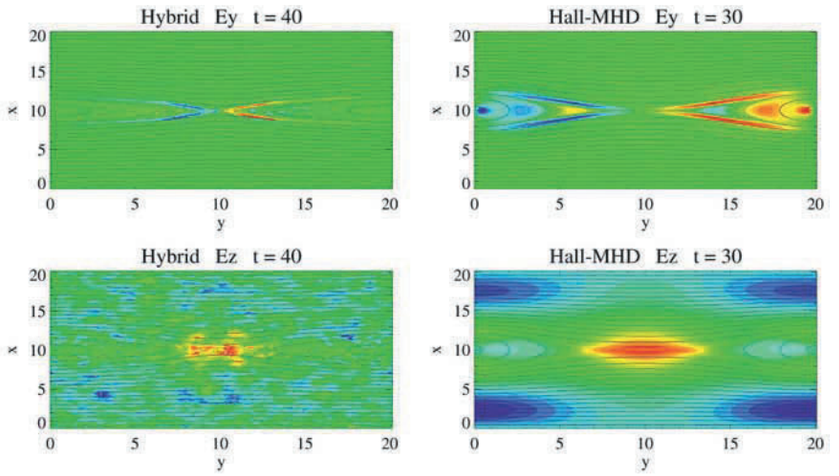


Fig. 7. Comparison of color contours of the electric field components E_y and E_z from hybrid (left panels) and Hall-MHD (right panels) simulations at comparable times.

run due to the absence of noise, the hybrid results at $t\Omega_i = 40$ are compared with the Hall-MHD results at $t\Omega_i = 30$. The color legend is the same as in Fig. 6, i.e., E_y and E_z have positive values in the red regions and negative values in the blue regions. General agreement is seen in the two calculations. Both runs show that the E_y component pointing away from the X point is enhanced in thin localized regions. These regions of strong E_y diverge from the X point and then follow the magnetic field lines at the edge of the sheet. The out-of-plane electric field E_z is significant mainly in the near-X-point region. The term $-\mathbf{V_e} \times \mathbf{B}$ in Eq. (31) provides the main contribution to the global electric fields, except at the X point where the magnetic field vanishes. Eq. (31) indicates that at the X point spatial gradients of the off-diagonal electron pressure tensor terms $-(\partial P^{e}_{\mathrm{xz}}/\partial x + \partial P^{e}_{\mathrm{yz}}/\partial y)$ provide a finite contribution to the E_z field (note that both η and m_e are set to zero). Thus this collisionless electron viscous effect plays the role of the conventional localized resistivity at the X point.

Overall agreement between the hybrid and Hall-MHD calculations is found from results of the 2-D reconnection problem in the reconnection rate, the global configuration of the currents and the fields, and the properties of the electron pressure [58]. However, in the Hall-MHD code ion dynamics are described by the ion momentum equation with a scalar ion pressure. Thus, ion finite Larmor radius effects contained in the off-diagonal terms of the ion pressure tensor are not modeled in the Hall-MHD code. This results in a significant difference in the spatial configuration of the out-of-plane ion velocity v_z from the two types of simulations as illustrated in Fig. 8: The v_z structure from the hybrid simulation (bottom panel) shows a reduction of the ion z-momentum at the X point as reconnection proceeds; in contrast, v_z from the Hall-MHD run (top panel) is peaked at the X point at this time ($t\Omega_i = 20$) and remains so peaked to the end of the run. It is shown that the ion finite Larmor radius effects are important to correctly model the ion out-of-plane momentum transport from the X point and can be modeled efficiently in Hall-MHD simulations in a predictor/corrector manner that uses particle ions to implement the ion gyro-radius corrections [59][60].

2.4 Global Hybrid Simulations

As stated in previous sections of this paper and illustrated in the references, hybrid simulations have been conducted and shown to be extremely beneficial in theoretical investigations of the bow shock, foreshock, magnetosheath, magnetopause, slow shocks in the magnetotail and wave particle interactions in the plasma sheet boundary layer. An overwhelming majority of these studies have been in the form of 1-D or planar 2-D simulations with specific emphasis on one of these topics. The results have clearly established the significance of ion kinetic physics in the temporal and spatial structure of these regions of geospace. By noting that these regions fall in the outer part of the magnetosphere (i.e. $R > 5 - 6R_E$) and that they are not heavily influenced by the ionospheric interactions or plasmas, we can construct a global model that would allow for a more comprehensive investigation of these regions. Specifically, this would allow

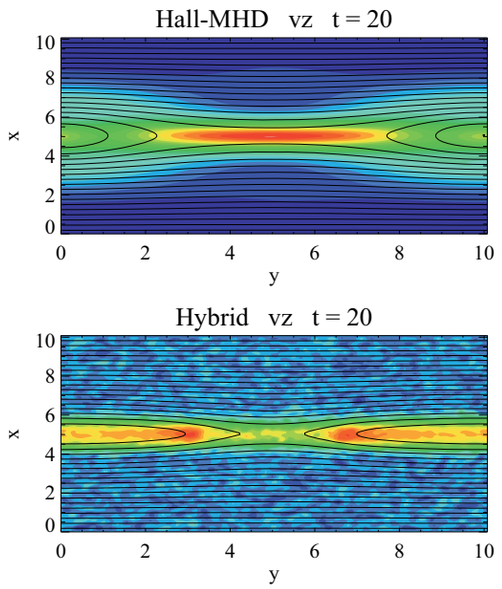


Fig. 8. Color contours of the ion z-velocity component at comparable times in Hall-MHD (top panel) and hybrid (bottom panel) simulations, showing differences at the X-point due to finite ion-gyroradius effects.

taking into account curvature and nonlocal effects as well as coupling between various parts of the outer magnetosphere (e.g., see [16]). While two- and three-dimensional global hybrid simulations have been performed to investigate the interaction between solar wind and Venus or Mars [61]–[63], the interaction is primarily with the respective ionospheres, where solar wind massloading is the dominant process. Also, the size of the interaction region is considerably smaller in comparison to the size of the Earth's magnetosphere, thereby making the CPU and memory requirements more manageable, even for computers of a decade ago. In contrast to Venus and Mars, a full investigation of the Earth's magnetosphere requires the inclusion of both its intrinsic magnetic field as well as the ionosphere. This, plus the large size of the magnetosphere, makes global hybrid simulations of this system considerably more complex and computationally more demanding.

A schematic of the global solar-wind magnetosphere simulation model is illustrated in Fig. 9. The hybrid simulations described here are two-dimensional in space, while all three components of the electromagnetic fields and ion velocities are retained [16][64]. Both the predictor-corrector field solver and ion velocity extrapolation methods described in subsection 2.B have been implemented, although the latter method is used commonly to reduce computation time. Due to the 2-D nature of the calculations, the Earth's magnetic field is represented by a line-dipole (see, e.g., [65]), which is placed within the simulation box at an arbitrary point. This dipole forms the center of a circular (Inner) region, which

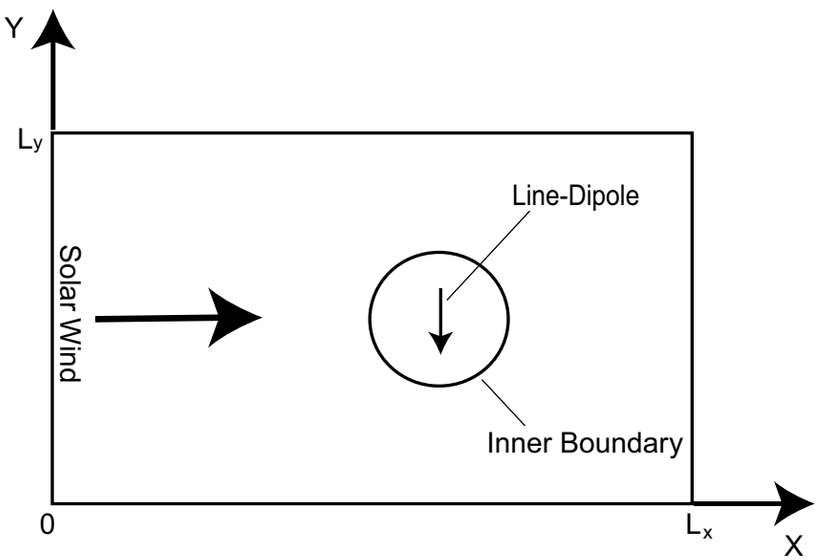


Fig. 9. The 2-D model used for global simulations. The solar wind is injected from the left boundary continuously during the run. The Earth's magnetic field is represented by a line-dipole that is placed at the center of a circular region which falls outside of the simulation domain.

falls outside of the simulation domain in that all particles crossing the "Inner Boundary" are lost and the electromagnetic fields are not solved for. This region represents the inner part of the magnetosphere that the current model does not appropriately describe; a similar approach is also taken in MHD simulations. Within the Inner Region, the electric field is set to zero and the magnetic field is kept at a constant level, which either corresponds to the dipole field strength at the Inner Boundary or is simply set to zero. More sophisticated sets of boundary conditions for the particles and the electromagnetic fields at the Inner Boundary could be applied to model magnetosphere-ionosphere coupling. The solar wind plasma is initially distributed within the simulation box and is also continuously injected from the $X = 0$ (left) boundary. This plasma is allowed to leave the system from all three remaining boundaries. The interplanetary magnetic field (IMF) is either in X-Y or X-Z plane making an arbitrary angle with the X-axis. The electric field at the $X = 0$ boundary is set to the value of the motional electric field ($\mathbf{V} \times \mathbf{B}$) in the solar wind. The remaining three boundaries have open or floating field boundary conditions in order to allow various waves and disturbances to leave the system. Specifically, the normal component of the electric field is set to zero on these boundaries, while the transverse components of the electric field are continuous across them. Typically, the cell size is the proton inertial length in the solar wind, which corresponds to ~ 100 km. The choice of time step varies from run to run depending on the solar wind conditions and the

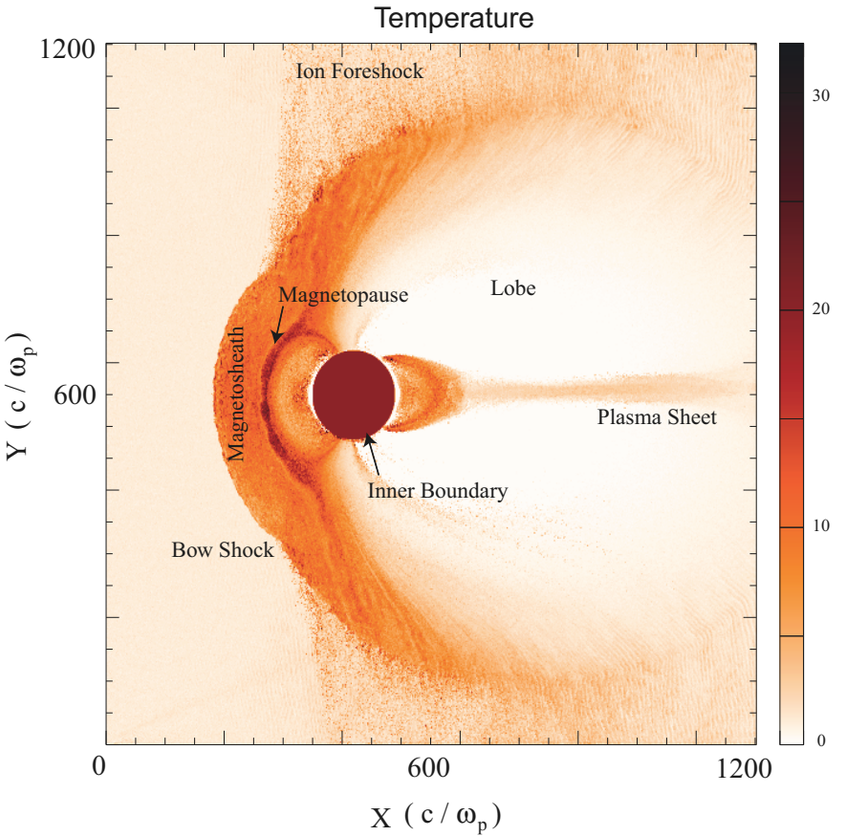


Fig. 10. Ion temperature as a function of X and Y in a representative global hybrid simulation. As denoted, regions representing various parts of the outer magnetosphere are formed in the simulations. The size of the simulation is roughly $19R_E \times 19R_E$.

dipole field strength. Typically, however, the time steps needed for numerical stability are much smaller than what one might estimate based on cell size and solar wind velocity.

Although initially the plasma is uniformly distributed in the simulation box, the evolution of the system in time results in the formation of regions and boundaries representing various parts of the outer magnetosphere. This is illustrated in Fig. 10, which shows the ion temperature, as a function of X and Y, for a representative run. In this run, the IMF lies in the X-Y plane and is almost perpendicular to the X-axis ($\theta = 85^o$). The figure shows the formation of the bow shock, ion foreshock, magnetosheath, magnetopause, the lobe and the plasma sheet. In this run, the simulation box is 1200 ion inertial lengths (in the solar wind) in X and Y directions and consists of 1,440,000 cells. Initially, each cell has 9 macroparticles but this number changes as various parts of the magnetosphere

with corresponding different plasma densities are formed. Assuming the ion inertial length to be 100 km, the size of the simulation box is about $19R_E \times 19R_E$, which is clearly smaller than the actual size of the magnetosphere. Similarly, because of the 2-D nature of the simulations, the system does not necessarily reach a steady state solution. This implies that these simulations cannot be used for determining the standoff position of the various discontinuities in the magnetosphere. On the other hand, the physical scale lengths associated with these boundaries (e.g. the thickness of the bow shock or the magnetopause) are appropriately modeled [16][64] and can be compared directly with spacecraft observations. Similarly, many details of the solar wind interaction with the magnetosphere that require a kinetic treatment, such as transport at the dayside magnetosphere, are properly modeled in these simulations.

Figure 10 clearly illustrates that the simulated magnetosphere bears considerable resemblance to the actual magnetosphere. This similarity goes beyond simple appearances; examination of plasma and field properties within the simulated magnetosphere show quantitative agreement with the spacecraft observations. For example, examination of the magnetic structure of the simulated bow shock in the quasi-perpendicular geometry shows considerable similarities to that of 1-D hybrid simulations as well as magnetometer data (e.g., see [19]). Similarly, evolution of the ion distribution functions across the shock reproduce the major features expected based on theory and observations [66]. This is demonstrated in Fig. 11, which shows the ion distribution functions in the upstream (top left), within the shock layer (top and bottom right) and the downstream regions (bottom left). The distribution functions within the shock layer show the characteristic presence of the gyrating ions which come about as part of the dissipation process (e.g., [8]). Evidently, the detailed study of such features in the simulations, coupled with more highly resolved observations through clusters of spacecraft, will transform our view of the dynamic magnetosphere.

The similarity between the simulated and actual magnetosphere provides a strong argument for the general validity of the model described here and shows the considerable potential it offers in understanding the dynamic and complex behavior of the magnetosphere under various solar wind conditions. This general agreement, however, is not a guarantee that numerical artifacts do not affect the specifics of a particular run, and the daunting task of assuring the physical nature of the results is a major part of the simulator's overall activities.

In a given global hybrid simulation of the magnetosphere, the simulation domain consists of many distinct regions differing in magnetic and plasma parameters. The usual tests devised for simpler systems to assess the significance of numerical artifacts in a given simulation run (e.g., conservation of energy, Courant condition) are inadequate in the highly inhomogeneous but coupled magnetospheric system. For instance, it can occur that results in one region of the simulation domain are physical but the results in another region are not. This is because the plasma and field properties change considerably within the magnetosphere (both actual and simulated), and therefore numerical conditions that are satisfied in one region may not be satisfied in another. An example of

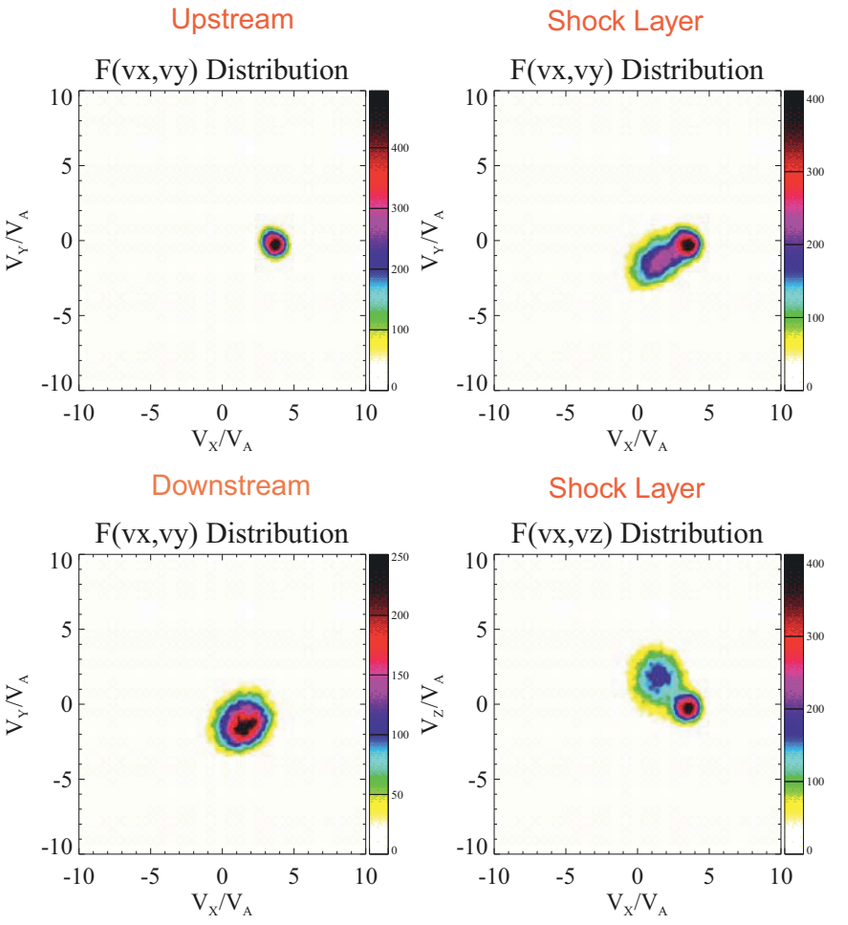


Fig. 11. Evolution of the ion distribution function across the quasi-perpendicular portion of the shock in a global simulation. This evolution is consistent with spacecraft observations at the bow shock and is similar to that seen in 1- and 2-D local hybrid simulations.

this is the drastic change in the plasma density from the magnetosheath into the lobe, which results in change in proton inertial length by an order of magnitude. In the case of the simulated lobe, the number of macroparticles in a cell may be one, or even less, requiring additional provisions to assure numerical stability. Since the hybrid algorithm does not contain electrostatic effects, the particle density in a given cell is never allowed to go below a certain base level. In the case of a simulated lobe, the presence of a cold, tenuous, and stationary plasma in the tail is assumed, much like the real magnetosphere; however, the plasma density in a cell is not allowed to fall below 5% of the solar wind density. Given the complexities involved, the best way of assuring the physical nature of the

results is through conducting many test runs and detailed examination of the results. This makes the process quite time consuming, but fortunately with the ever-increasing speed of the computers, this task is becoming more and more manageable.

3 Hybrid Codes: Future

In this tutorial, we also speculate on the future development and use of hybrid codes for space physics applications, as is already evident to some degree from the many papers at ISSS-6. We can see significant progress occurring in five general areas [12][19][67]: (1) larger and more complex simulations, (2) inclusion of more physics, (3) improvements in diagnostics for better physical insight and comparison with data, (4) algorithm development for massively parallel computers, and (5) linking hybrid and fluid codes together. Some examples of recent past progress and future work in these areas include the following.

(1) The availability of faster CPUs, more memory, etc. will lead to larger scale simulations. Such calculations will include larger regions of space, e.g., the dayside magnetosphere [16][64], the magnetopause [17][68]–[70] and the magnetotail [50][54][71][72]. In addition, three-dimensional effects, and/or more complex multi-species problems can be investigated, such as the solar wind interaction with comets [14][73][74] unmagnetized planets [61]–[63], asteroids, etc., as well as modeling kinetic processes in the expanding solar wind [75], beyond that which is presently available. We have already shown the potential, as well as some of the difficulties, for improved understanding of nonlocal processes through global hybrid simulations in the previous section.

(2) More complex physics models may include, for example, semi-collisional plasmas, such as occur in the polar region, where the outflowing plasma is collisional near the Earth and becomes less so as it flows outward [76] and at comets. This can involve new types of collision models, such as collision-field methods in which the collision "force" is treated as a grid quantity [77][78]. We have already discussed another possibility: namely, the use of hybrid codes to understand new electron and ion kinetic effects that occur near the reconnection site, which can be modeled in Hall-MHD and MHD codes.

(3) Improvements in diagnostics are likely to come through the use of commercial products, like IDL or EnSight, as the development of major visualization tools, especially in 3-D, is far too expensive for any particular research group. Hybrid codes offer unique possibilities for development of diagnostics that examine ion distribution functions, ion-scale fluctuations, etc., which can be expressed in a form convenient for comparing with data. This will be particularly useful for understanding spatial and temporal correlations between data from several different satellites, e.g., Cluster II, and global simulations.

(4) Computers that consist of 1000's of linked processors seem to be the most economical pathway for large scale computing. Again, these can be very expensive machines that only the largest institutions can afford or they can be a group of inexpensive PC's or Mac's that are ganged together [79]. Understanding how

to write algorithms that take advance the particular system's unique architecture can be time-consuming, but can pay off in the long run. The challenge for hybrid (and other PIC) codes on massively parallel architectures is to balance the load between processors for pushing the particles and to break up the computational domain in a convenient manner [80]. For smaller scale problems, one can use the various processors to push the particles, solve the field equations over the entire domain on one processor, and distribute the field information to the other processors [27] in a very efficient manner.

(5) Finally, there is the issue of including kinetic physics in large-scale fluid calculations for developing realistic space weather codes. The required speed of such predictive codes precludes just doing global hybrid simulations [16][64]. As we have discussed earlier, kinetic effects found from hybrid simulations can be modeled in Hall-MHD codes [58][59][60]. In turn, a Hall-MHD code can be embedded in a global MHD code [60][81], thus providing an efficient way to include kinetic effects in a large-scale fluid code. Embedding an actual hybrid simulation in an MHD code would seem to be much more complex, given the disparate time and spatial scales between an ion kinetic model and an MHD model. In principle, this might be done be running the calculations concurrently on separate machines and exchanging appropriate boundary information to initialize the hybrid calculation or to update the MHD simulation each time step. It certainly provides the ultimate "grand-challenge" problem for graduates of ISSS-6!

Acknowledgements

At Los Alamos, this work was performed under the auspices of the Department of Energy and was supported in part by the NASA Sun Earth Connection Theory Program. Work at UCSD was performed under the auspices of California Space Institute and was supported by NSF grants ATM-9901665, ATM-0119846 and an IGPP grant from LANL. Support by the San Diego Supercomputer Center is also acknowledged.

References

1. Auer, P.L., Kilb, R.W., Crevier, W.F. (1971): Thermalization in the Earth's bow shock. J. Geophys. Res., **76**, 2927
2. Forslund, D.W., Freidberg, J.P. (1971): Theory of laminar collisionless shocks. Phys. Rev. Lett., **27**, 1189
3. Mason, R.J. (1971): Computer simulation of ion-acoustic shocks: The diaphragm problem. Phys. Fluids, **14**, 1943
4. Chodura, R. (1975): A hybrid fluid-particle model of ion heating in high-Mach-number shock waves. Nucl. Fusion, **15**, 55
5. Sgro, A.G., Nielson, C.W. (1976): Hybrid model studies of ion dynamics and magnetic field diffusion during pinch implosions. Phys. Fluids, **19**, 126
6. Hamasaki, S., Krall, N.A., Wagner, C.E., Byrne, R.N. (1977): Effects of turbulence on theta pinch modeling by hybrid numerical methods. Phys. Fluids, **20**, 65
7. Hewett, D.W. (1980): A global method for solving the electron-field equations in a zero-inertial electron hybrid plasma simulation code. J. Comput. Phys., **38**, 378

8. Leroy, M.M., Goodrich, C.C., Winske, D., Wu, C.S., Papadopoulos, K. (1981): Simulation of a perpendicular bow shock. Geophys. Res. Lett., **8**, 1269
9. Winske, D., Leroy, M.M. (1984): Hybrid simulation techniques applied to the Earth's bow shock. In: Matsumoto, H., Sato, T. (ed) Computer Simulations of Space Plasmas. D. Reidel/Terra Sci., Hingham MA
10. Winske, D. (1985): Hybrid simulation codes with application to shocks and upstream waves. Space Sci. Rev., **42**, 53
11. Quest, K.B. (1989): Hybrid simulation. In: Lembege, B., Eastwood, J. W. (ed) Tutorial Courses: Third International School for Space Simulation. Cepadues Ed., Toulouse, France
12. Winske, D., Omidi, N. (1993): Hybrid codes: Methods and applications. In: Matsumoto, H., Omura, Y. (ed) Computer Space Plasma Physics: Simulations and Software. Terra, Tokyo
13. Pritchett, P.L. (2002): Particle-in-cell simulations of plasmas. This volume
14. Bagdonat, T., Motschmann, U. (2001): 3-D hybrid simulation of the solar wind interaction with comets. In: Buechner, J., Dum, C.T., Scholer, M. (ed) Space Plasma Simulation: Proc. Sixth International School/Symposium ISSS-6. Copernicus Gesellschaft, Katlenburg-Lindau, Germany
15. Lipatov, A.S. (2001): Hybrid codes with finite electron mass. In: Buechner, J., Dum, C.T., Scholer, M. (ed) Space Plasma Simulation: Proc. Sixth International School/Symposium ISSS-6. Copernicus Gesellschaft, Katlenburg-Lindau, Germany
16. Omidi, N., Karimabadi, H., Quest, K.B. (2001): Global hybrid simulations of solar wind interaction with the magnetosphere. In: Buechner, J., Dum, C.T., Scholer, M. (ed) Space Plasma Simulation: Proc. Sixth International School/Symposium ISSS-6. Copernicus Gesellschaft, Katlenburg-Lindau, Germany
17. Quest, K.B., Karimabadi, H. (2001): Physics of solar wind magnetosphere interactions: 3-D hybrid simulations. In: Buechner, J., Dum, C.T., Scholer, M. (ed) Space Plasma Simulation: Proc. Sixth International School/Symposium ISSS-6. Copernicus Gesellschaft, Katlenburg-Lindau, Germany
18. Yin, L., Winske, D., Gary, S.P., Birn, J. (2001): Hybrid and Hall-MHD simulations of collisionless reconnection: Effects of plasma pressure tensor. In: Buechner, J., Dum, C.T., Scholer, M. (ed) Space Plasma Simulation: Proc. Sixth International School/Symposium ISSS-6. Copernicus Gesellschaft, Katlenburg-Lindau, Germany
19. Winske, D., Omidi, N. (1996): A nonspecialist's guide to kinetic simulations of space plasmas. J. Geophys. Res., **101**, 17287
20. Pritchett, P.L. (2000): Particle-in-cell simulations of magnetosphere electrodynamics. IEEE Trans. Plasma Sci., **28**, 1976
21. Omura, Y., Huba, J.D., Winske, D. (1999): Theory and simulations of nonlinear kinetic processes in space plasmas. In: Stone, W.R. (ed) The Review of Radio Science 1996–1999. Oxford Univ. Press, Oxford
22. Lipatov, A.S. (2001): The Hybrid Multiscale Simulation Technology. Springer-Verlag, Heidelberg, Germany
23. Vu, H.X., Brackbill, J.U. (1992): CELEST1D: An implicit fully kinetic model for low frequency electromagnetic plasma simulation. Comput. Phys. Comm., **69**, 253
24. Oppenheim, M., Otani, N., Ronchi, C. (1996): Saturation of the Farley-Buneman instability via nonlinear electron $E \times B$ drifts. J. Geophys. Res., **101**, 17273
25. Mankofsky, A., Sudan, R.N., Denavit, J. (1987): Hybrid simulation of ion beams in background plasma. J. Comput. Phys., **70**, 89
26. Omelchenko, Yu.A., Sudan, R.N. (1997): A 3-D Darwin-EM hybrid PIC code for ion ring studies. J. Comput. Phys., **133**, 146

27. Vu, H.X. (1998): A massively parallel three-dimensional hybrid code for simulating ion-driven parametric instabilities. J. Comput. Phys., **144**, 257
28. Birdsall, C.K., Langdon, A.B. (1985): Plasma Physics Via Computer Simulation. McGraw-Hill, New York
29. Swift, D.W. (1995): Use of a hybrid code to model the Earth's magnetosphere. Geophys. Res. Lett., **22**, 311
30. Byers, J.A., Cohen, B.I., Condit, W.C., Hanson, J.D. (1978): Hybrid simulations of quasi-neutral phenomena in magnetized plasmas. J. Comput. Phys., **27**, 363
31. Harned, D.S. (1982): Quasineutral hybrid simulation of macroscopic plasma phenomena. J. Comput. Phys., **47**, 452
32. Winske, D., Quest, K.B. (1986): Electromagnetic ion beam instabilities: Comparison of one- and two-dimensional simulations. J. Geophys. Res., **91**, 8789
33. Brecht, S.H., Thomas, V.A. (1988): Multidimensional simulations using hybrid particle codes. Computer Phys. Comm., **48**, 135
34. Fujimoto, M. (1991): Instabilities in the magnetopause velocity shear layer. Thesis, U. Tokyo
35. Thomas, V.A., Winske, D., Omidi, N. (1990): Reforming supercritical quasi-parallel shocks: 1. One- and two-dimensional simulations. J. Geophys. Res., **95**, 18809
36. Winske, D., Quest, K.B. (1988): Magnetic field and density fluctuations at perpendicular supercritical collisionless shocks. J. Geophys. Res., **93**, 9681
37. Matthews, A.P. (1994): Current advance method and cyclic leapfrog for 2-D multispecies hybrid plasma simulations. J. Comput. Phys., **112**, 102
38. Horowitz, E.J., Shumaker, D.E., Anderson, D.V. (1989): QN3D: A three-dimensional quasi-neutral hybrid particle-in-cell code with applications to the tilt instability in field reversed configurations. J. Comput. Phys., **84**, 279
39. Terasawa, T., Hoshino, M., Sakai, J.I., Hada, T. (1986): Decay instability of finite amplitude circularly polarized Alfvénwaves: A numerical simulation of stimulated Brillouin scattering. J. Geophys. Res., **91**, 4171
40. Kazeminezhad, F., Dawson, J.M., Leboeuf, J.N., Sydora, R., Holland, D. (1992): A Vlasov particle ion zero mass electron model for plasma simulations. J. Comput. Phys., **102**, 277
41. Hewett, D.W., Nielson, C.W. (1978): A multidimensional quasineutral plasma simulation model. J. Comput. Phys., **29**, 219
42. Shay, M., Drake J.F., Denton, R.E., Biskamp, D. (1998): Structure of the dissipation region during collisionless magnetic reconnection. J. Geophys. Res., **103**, 9165
43. Kuznetsova, M.M., Hesse, M., Winske, D. (1998): Kinetic quasi-viscous and bulk flow inertia effects in collisionless magnetotail reconnection. J. Geophys. Res., **103**, 199
44. Omidi, N., Winske, D. (1992): Kinetic structure of slow shocks: Effects of the electromagnetic ion/ion cyclotron instability. J. Geophys. Res., **97**, 14801
45. Winske, D., Omidi, N. (1992): Electromagnetic ion/ion cyclotron instability: Theory and simulations. J. Geophys. Res., **97**, 14779
46. Cremer, M., Scholer, M. (1999): Collisionless slow shocks in magnetotail reconnection. Geophys. Res. Lett., **26**, 2709
47. Cremer, M., Scholer, M. (2000): Structure of the reconnection layer and the associated slow shocks: Two-dimensional simulations of a Riemann problem. J. Geophys. Res., **105**, 27621
48. Daughton, W., Gary, S.P., Winske, D. (1999): Electromagnetic proton/proton instabilities in the solar wind: Simulations. J. Geophys. Res., 104, 4657

49. Nakamura, M.S., Fujimoto, M. (1998): A three-dimensional hybrid simulation of magnetic reconnection. Geophys. Res. Lett., **25**, 2917
50. Krauss-Varban, D., Omidi, N. (1995): Large-scale hybrid simulations of the magnetotail during reconnection. Geophys. Res. Lett., **22**, 3271
51. Kuznetsova, M.M., Hesse, M., Winske, D. (1996): Ion dynamics in a hybrid simulation of magnetotail reconnection. J. Geophys. Res., **101**, 27351
52. Nakabayashi, J., Machida, S. (1997): Electromagnetic hybrid code simulation of magnetic reconnection: Velocity distribution functions of accelerated ions. Geophys. Res. Lett., **24**, 1339
53. Nakamura, M.S., Fujimoto, M. (1998): Ion dynamics and resultant velocity space distributions in the course of magnetotail reconnection. J. Geophys. Res., **103**, 4531
54. Lottermoser, R.F., Scholer, M., Matthews, A.P. (1998): Ion kinetics in magnetic reconnection: Hybrid simulations. J. Geophys. Res., **103**, 4547
55. Hesse, M., Winske, D. (1993): Hybrid simulations of collisionless ion tearing. Geophys. Res. Lett., **20**, 1207
56. Hesse, M., Winske, D. (1994): Hybrid simulations of collisionless reconnection in current sheets. J. Geophys. Res., **99**, 11177
57. Hesse, M., Winske, D., Kuznetsova, M.M. (1995): Hybrid modeling of collisionless reconnection in two-dimensional current sheets: Simulations. J. Geophys. Res., **100**, 21815
58. Yin, L., Winske, D., Gary, S.P., Birn, J. (2001): Hybrid and Hall-MHD simulations of collisionless reconnection: Dynamics of the electron pressure tensor. J. Geophys. Res., **106**, 10761
59. Yin, L., Winske, D., Gary, S.P., Birn, J. (2001): Particle Hall-MHD simulation of collisionless reconnection: Finite ion gyro-orbit correction. Geophys. Res. Lett., **28**, 2173
60. Yin, L., Winske, D., Gary, S.P., Birn, J. (2002): Hybrid and Hall-magnetohydrodynamics simulations of collisionless reconnection: Effect of the ion pressure tensor and particle Hall-magnetohydrodynamics. Phys. Plasmas, **9**, 2575-2584
61. Brecht, S.H., Ferrante, J.R. (1991): Global hybrid simulation of unmagnetized planets: Comparison of Venus and Mars. J. Geophys. Res., **96**, 11209
62. Brecht, S.H., Ferrante, J.R., Luhmann, J.G. (1993): Three-dimensional simulations of the solar wind interaction with Mars. J. Geophys. Res., **98**, 1345
63. Brecht, S.H. (1997): Hybrid simulations of the magnetic topology of Mars. J. Geophys. Res., **102**, 4743
64. Omidi, N., Karimabadi, H., Krauss-Varban, D¿ (1998): Hybrid simulation of the curved dayside magnetopause during southward IMF. Geophys. Res. Lett., **25**, 3273
65. Ogino, T. (1993): Two-dimensional MHD code. In: Matsumoto, H., Omura, Y. (ed) Computer Space Plasma Physics: Simulations and Software. Terra, Tokyo
66. Sckopke, N., Paschmann, G., Bame, S.J., Gosling, J.T., Russell, C.T. (1983): Evolution of ion distributions across the nearly perpendicular bow shock: Specularly and non-specularly reflected-gyrating ions. J. Geophys. Res., **88**, 6121
67. Dawson, J.M. (1999): Role of computer modeling of plasmas in the 21st century. Phys. Plasmas, **6**, 4436
68. Lin, Y., Xie, H. (1997): Formation of reconnection layer at the dayside magnetopause. Geophys. Res. Lett., **24**, 3145
69. Xie, H., Lin, Y. (2000): Two-dimensional hybrid simulation of the dayside reconnection layer and associated ion transport. J. Geophys. Res., **105**, 25171

70. Nakamura, M., Scholer, M. (2000): Structure of the magnetopause reconnection layer and of flux transfer events: Ion kinetic effects. J. Geophys. Res., **105**, 23179
71. Lin, Y., Swift, D.W. (1996): A two-dimensional hybrid simulation of the magnetotail reconnection layer. J. Geophys. Res., **101**, 19859
72. Scholer, M., Lottermoser, R.F. (1998): On the kinetic structure of the magnetotail reconnection layer. Geophys. Res. Lett., **25**, 3281
73. Galeev, A.A., Lipatov, A.S., Sagdeev, R.Z. (1987): Two-dimensional numerical simulation of the relaxation of cometary ions and MHD turbulence in the flow of the solar wind around a cometary atmosphere. Sov. J. Plasma Phys., **13**, 323
74. Omidi, N., Winske, D. (1990): Steepening of magnetosonic waves into shocklets: Simulations and consequences for planetary shocks and comets. J. Geophys. Res., **95**, 2281
75. Liewer, P.C., Velli, M., Goldstein, B. (1999): Hybrid simulations of wave propagation and ion cyclotron heating in the expanding solar wind. Space Sci. Rev., **87**, 257
76. Miller, R.H., Rasmussen, C.E., Gombosi, T.I., Khazanov, G.V., Winske, D. (1993): Kinetic simulation of plasma flows in the inner magnetosphere. J. Geophys. Res., **98**, 19301
77. Jones, M.E., Lemons, D.S., Mason, R.J., Thomas, V.A., Winske, D. (1996): A grid-based Coulomb collision model for PIC codes. J. Comput. Phys., **123**, 169
78. Thomas, V.A. (1997): Modeling combined collisional/collisionless plasma interpenetration. IEEE Trans. Plasma Sci., **25**, 353
79. Decyk, V.K., Dauger, D.E. (2001): How to build an AppleSeed: A parallel Macintosh cluster for numerically intensive computing. In: Buechner, J., Dum, C.T., Scholer, M. (ed) Space Plasma Simulation: Proc. Sixth International School/Symposium ISSS-6. Copernicus Gesellschaft, Katlenburg- Lindau, Germany
80. Liewer, P.C., Decyk, V.K. (1989): A general concurrent algorithm for plasma particle-in-cell simulation codes. J. Comput. Phys., **85**, 302
81. Omidi, N., Blanco-Cano, X., Russell, C.T., Karimabadi, H., Acuna, M. (2002): Hybrid simulations of solar wind interaction with magnetized asteroids: General characteristics. J. Geophys. Res., **107**, in press

Hall Magnetohydrodynamics - A Tutorial

Joseph D. Huba

Plasma Physics Division, Naval Research Laboratory, Washington, DC 20375

Abstract. Over the past fifteen years it has become increasingly clear that Hall magnetohydrodynamics plays a crucial role in many space and laboratory plasma processes: magnetic reconnection, sub-Alfénic plasma expansions, and plasma opening switches to name a few. Hall magnetohydrodynamics is important for plasma dynamics on length scales less than the ion inertial scale length but greater than the electron inertial length. On these scales the ion and electron motions are decoupled; the electrons remain frozen to the magnetic field but the ions are not. In this paper we provide a basic overview of Hall magnetodydrodynamics with an emphasis on numerical methods. We also provide several concrete examples of Hall dynamics: whistler waves, Hall drift waves, plasma opening switch dynamics, and three dimensional magnetic reconnection.

1 Introduction

Hall magnetohydrodynamic (MHD) theory has been used to describe and understand a variety of interesting space and laboratory plasma phenomena over the last two decades. The theory has been successfully applied to the structuring of sub-Alfvénic plasma expansions [1,2,3,4,5,6,7] and to rapid magnetic field transport in plasma opening switches [8,9,10,11,12,13,14,15,16,17]. In recent years it has become evident that Hall physics plays a critical role in magnetic reconnection processes [18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36] and this has spurred renewed interest in this subject. Generally speaking, the theory is applicable to phenomena occurring on length scales shorter than an ion inertial length, and time scales shorter than an ion cyclotron period. In this paper we will provide an overview of the underlying physics associated with Hall physics, discuss numerical techniques to incorporate the Hall term into an MHD code, and describe several applications of Hall MHD to space and laboratory plasmas.

2 Hall MHD: Basic Equations and Wave Modes

The essence of Hall MHD physics is contained in Ohm's law. The generalized form of Ohm's law can be written as [37]

$$\frac{m_e}{ne^2}\frac{\partial \mathbf{J}}{\partial t} - \frac{1}{ne}\nabla\cdot\overline{\overline{P}}_e = \mathbf{E} + \frac{1}{c}\mathbf{V}\times\mathbf{B} - \frac{1}{nec}\mathbf{J}\times\mathbf{B} - \mathbf{J}/\sigma \tag{1}$$

where σ is the plasma conductivity. We simplify (1) by assuming $L >> c/\omega_{pe}$, $L >> \rho_e$, and $\sigma \to \infty$ where ω_{pe} is the electron plasma frequency, ρ_e is the

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 166–192, 2003.

electron Larmor radius, and L is the scale length of the plasma. These assumptions allow us to neglect the electron inertia and electron pressure terms, and the conductivity term. We then obtain

$$\mathbf{E} = -\frac{1}{c}\mathbf{V}_i \times \mathbf{B} + \underbrace{\frac{1}{nec}\mathbf{J} \times \mathbf{B}}_{Hall\ term} \ . \tag{2}$$

The final term in (2) is identified as the Hall term. Physically, the Hall term decouples ion and electron motion on ion inertial length scales: $L < c/\omega_{pi}$ where $\omega_{pi} = 4\pi n_i e^2/m_i$.

The full set of Hall MHD equations is as follows:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \mathbf{V} = 0 \tag{3}$$

$$\frac{\partial \rho \mathbf{V}}{\partial t} + \nabla \cdot \left[\rho \mathbf{V}\mathbf{V} + (P + B^2/8\pi)\underline{\underline{\mathbf{I}}} - \mathbf{B}\mathbf{B}/4\pi\right] = 0 \tag{4}$$

$$\frac{\partial \epsilon}{\partial t} + \nabla \cdot \left[\mathbf{V}(\epsilon + P + B^2/8\pi) - \mathbf{B}/4\pi(\mathbf{V} \cdot \mathbf{B})\right] + \nabla \cdot \left[\mathbf{V}_H(B^2/8\pi) - \mathbf{B}/4\pi(\mathbf{V}_H \cdot \mathbf{B})\right] = 0 \tag{5}$$

$$\frac{\partial \mathbf{B}}{\partial t} = -c\nabla \times \mathbf{E} = \nabla \times [(\mathbf{V} + \mathbf{V}_H) \times \mathbf{B}] \tag{6}$$

where $\epsilon = \rho V^2/2 + P/(\gamma - 1) + B^2/8\pi$, $P/\rho^\gamma = \text{cnst}$, and $\mathbf{V}_H = -\mathbf{J}/ne$. The plasma pressure is obtained from

$$P = (\gamma - 1)(\epsilon - \rho V^2/2 - B^2/8\pi) \ . \tag{7}$$

We take $\gamma = 5/3$. In writing (3) – (6) the subscript i has been dropped for simplicity, and the variable $\mathbf{V}_H$ is defined as a 'Hall velocity' to explicitly show where the Hall term enters the equations. Aside from the magnetic field induction equation (6), the Hall term only enters the energy equation (5). Thus, the Hall term is a transport mechanism for the magnetic field but not for mass or momentum.

We point out that solving the energy equation to calculate the pressure is problematic in low β plasmas because it involves subtracting two large numbers (i.e., ϵ and $B^2/8\pi$). An alternative to using the energy equation (5) is to use the adiabatic pressure equation

$$\frac{\partial P}{\partial t} + \nabla \cdot P\mathbf{V} = -(\gamma - 1)P\nabla \cdot \mathbf{V} \ . \tag{8}$$

However, a computational problem arises using (8): the Rankine-Hugoniot jump conditions across a shock are not correct unless additional modifications are made (e.g., inclusion of an artificial viscosity); whereas the jump conditions are automatically satisfied using the total energy equation.

The Hall term introduces two new wave modes into the plasma system: whistler waves and Hall drift waves. It is important that these wave modes be identified and understood in developing a Hall simulation code because they determine the time step through the Courant condition.

The dispersion relations for these wave modes are easily derived in the electron magnetohydrodynamic (EMHD) limit: $\mathbf{V} = 0$. In this limit the ions are assumed to be a stationary, neutralizing background. The EMHD magnetic induction equation is

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{V}_H \times \mathbf{B}) = -\nabla \times (\mathbf{J}/ne \times \mathbf{B}) \ . \tag{9}$$

We rewrite (9) as follows to identify the origin of each wave mode

$$\frac{\partial \mathbf{B}}{\partial t} \ = \ \underbrace{-\frac{1}{ne}\nabla \times (\mathbf{J} \times \mathbf{B})}_{\text{Whistler waves}} \ \underbrace{+ \frac{1}{n^2 e}\nabla n \times (\mathbf{J} \times \mathbf{B})}_{\text{Hall drift waves}} \ . \tag{10}$$

The first term on the RHS of (10) is responsible for the whistler wave, while the second term for the Hall drift wave. It is important to note that Hall drift waves only occur in inhomogeneous plasmas. We now discuss each wave mode in more detail.

2.1 Whistler Waves

We assume the following plasma configuration. The plasma is homogeneous and the ambient magnetic field is in the $z-$direction: $\mathbf{B} = B_0 \ \hat{\mathbf{e}}_z$. The magnetic field is perturbed with δB_x and $\delta B_y \propto \exp(ik_z z - i\omega t)$. The linear dispersion equation is obtained from

$$\frac{\partial \delta \mathbf{B}}{\partial t} \ = \ -\frac{1}{ne}\nabla \times (\delta \mathbf{J} \times \mathbf{B}) \ . \tag{11}$$

We linearize (11) to obtain

$$\omega \delta \mathbf{B} = \frac{1}{ne}(k_z B \delta J_x \ \hat{\mathbf{e}}_x + k_z B \delta J_y \ \hat{\mathbf{e}}_y) \ . \tag{12}$$

Making use of Ampere's law (i.e., $\delta \mathbf{J} = (c/4\pi)\nabla \times \delta \mathbf{B}$) we obtain the coupled equations

$$\omega \delta B_x = -i\frac{ck_z^2 B}{4\pi ne}\delta B_y \tag{13}$$

$$\omega \delta B_y = i\frac{ck_z^2 B}{4\pi ne}\delta B_x \ . \tag{14}$$

Finally, solving (13) and (14) the whistler wave dispersion relation is obtained

$$\omega = \frac{k_z^2 B}{4\pi ne} = k_z V_A \left(\frac{k_z c}{\omega_{pi}}\right) \ . \tag{15}$$

It is evident from the final form of the dispersion relation that the whistler wave phase velocity exceeds the Alfvén velocity for wavelengths $\lambda_z < c/\omega_{pi}$. This makes it clear that the Hall term is important for scale sizes less than the ion inertial length as noted in the Introduction.

2.2 Hall Drift Waves

We assume the following plasma configuration. The plasma density is inhomogeneous $n(x)$ and the ambient magnetic field is in the $z-$direction: $\mathbf{B} = B_0\ \hat{\mathbf{e}}_z$. The ambient magnetic field is perturbed with $\delta B_z \propto \exp(ik_y y - i\omega t)$. The linear dispersion equation is obtained from

$$\frac{\partial \delta \mathbf{B}}{\partial t} = \frac{1}{n^2 e} \nabla n \times (\delta \mathbf{J} \times \mathbf{B}) \ . \tag{16}$$

We linearize (16) and find that

$$i\omega \delta B_z = \frac{1}{n^2 e} \frac{\partial n}{\partial x} \delta J_x B \ . \tag{17}$$

Noting that $\delta J_x = ik_y (c/4\pi) \delta B_z$, the following dispersion relation is obtained from (17)

$$\omega = \frac{k_y B}{4\pi n e} \frac{1}{n} \frac{\partial n}{\partial x} = k_y V_A \left(\frac{c}{L_n \omega_{pi}} \right) \tag{18}$$

where $L_n = (\partial ln\, n/\partial x)^{-1}$ is the density gradient scale length. Again, it is clear that the Hall term is important when $L_n < c/\omega_{pi}$. The Hall drift wave described by (18) is a magnetic drift wave that propagates in the $\mathbf{B} \times \nabla n$ direction. A detailed analysis of this wave mode is presented in [11] and [38].

In addition to the propagation of a linear wave mode, it has also been shown that shock-like solutions exist based on a nonlinear analysis of (9) [39,40]. Recently, Rudakov and Huba [41] found that a rapid, localized thinning of a current layer supported by a uni-directional magnetic field leads to the generation of a nonlinear, shock-like structure. This shock-like wave also propagates in the $\mathbf{B} \times \nabla n$ direction; it is self-supportive and can lead to a nonlocal thinning of the current layer and the release of magnetic energy.

3 Numerical Methods

We describe in some detail the numerical techniques used in the NRL 3D Hall MHD code VooDoo to solve (3) – (6). The key points we address are the cell definition, the finite volume method, the time step scheme, the high order interpolation scheme, the partial donor cell flux limiter, the distribution function method, the calculation of the electric field, the Courant condition, and subcycling the Hall term for computational efficiency.

3.1 Cell Definition

The code uses a Cartesian, staggered mesh, known as the Yee grid, and is shown in Fig. 1. The hydrodynamic variables mass density ρ, velocity $\mathbf{V}$, and energy density ϵ are defined at the cell center. The magnetic field $\mathbf{B}$ is defined normal to the cell faces, and the electric field $\mathbf{E}$ is defined along the cell edges. The advantage to using this grid is that $\nabla \cdot \mathbf{B} = 0$ is satisfied to machine error.

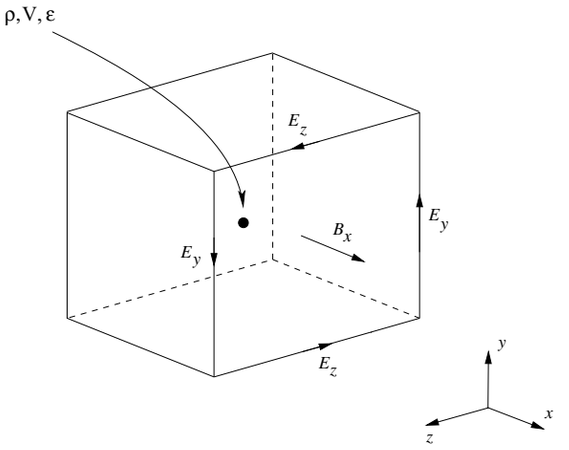


Fig. 1. Mesh used in the NRL 3D Hall code VooDoo.

3.2 Time Step Scheme

The code uses a 2nd order Adams-Bashforth time stepping scheme. The time advancement scheme to update the variable A from time t to time $t + \Delta t$ is

$$A^{t+\Delta t} = A^t + \Delta t \mathcal{F}_A^{t+\Delta t/2} \tag{19}$$

where $\mathcal{F}_A^{t+\Delta t/2}$ is the total flux of A through the cell faces at the half time step. The values of the variables at the half time step needed to calculate the flux F are obtained by a linear extrapolation from the two previous time steps. Specifically, the half time step values are given by

$$A^{t+\Delta t/2} = A^t + \frac{\Delta t}{2\Delta t_0}\left(A^t - A^{t-\Delta t_0}\right) \tag{20}$$

where the time steps are shown schematically in Fig. 2. The code uses a variable time step so the full time step increments (Δt_0 and Δt) can be different.

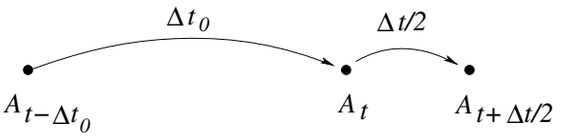


Fig. 2. Schematic of the time step scheme to obtain variables at the half time step.

3.3 Finite Volume Method

The finite volume method updates the conserved variables mass, momentum, and energy by calculating the fluxes of these variables across the cell faces. This

is more clearly understood by considering an example. The continuity equation is given by

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \rho \mathbf{V} \ . \tag{21}$$

This equation is integrated over a cell volume

$$\int \frac{\partial \rho}{\partial t} \ d^3x = -\int \nabla \cdot \rho \mathbf{V} \ d^3x \tag{22}$$

which can be written as

$$\frac{\partial \rho_T}{\partial t} = -\oint (\rho \mathbf{V}) \cdot \hat{\mathbf{n}} \ d^2x \tag{23}$$

where $\rho_T = \int \rho \, d^3x$ is the mass in the cell, $\hat{\mathbf{n}}$ is the normal to the cell face, and we have used Gauss' law. The total mass is then advanced in time from t to $t + \Delta t$ by

$$\rho_T^{t+\Delta t} = \rho_T^t - \Delta t \oint \underbrace{(\rho \mathbf{V})^{t+\Delta t/2}}_{\text{flux } \mathbf{F}} \cdot \hat{\mathbf{n}} \ d^2x \tag{24}$$

where we have explicitly identified the mass flux term. The density at the updated time step is then $\rho^{t+\Delta t} = \rho_T^{t+\Delta t} / \int d^3x$. This same technique is also used to update the momentum and energy. Thus, the code must calculate the mass, momentum, and energy fluxes at each cell face. The general updating scheme for A is

$$A^{t+\Delta t} = A^t - \int \mathbf{F}_A^{t+\Delta t/2} \cdot \mathbf{n} \ d^2x \tag{25}$$

where $\mathbf{F}_A$ is the flux of A.

3.4 Flux Calculation

The code VooDoo uses a distribution function scheme to calculate the fluxes of mass, momentum, and energy at cell faces. The details of this method are described in [42] and will not be repeated here. Rather, we will present an overview of the method, and descriptions of the high order interpolation scheme and the partial donor cell flux limiter.

The calculation of the flux at a cell face is as follows; we use the one dimensional continuity equation as an example. The mass flux is defined as

$$F = \rho V = \rho \int dv \, v \, f(v) \tag{26}$$

where $f(v)$ is a distribution function that describes the plasma system (this will be discussed shortly). We need to calculate this quantity at a cell face $i + 1/2$. The total flux F is decomposed into two parts, a left flux F_L and a right flux F_R, as follows

$$F_{i+1/2} = (\rho V)_{i+1/2} = (\rho V)^L_{i+1/2} + (\rho V)^R_{i+1/2} \tag{27}$$

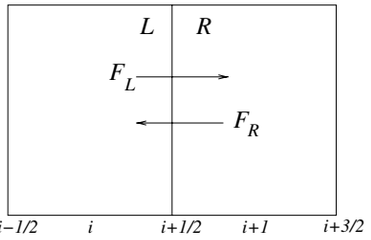


Fig. 3. Flux components F_L and F_R at a cell face $i + 1/2$.

where $F_L = (\rho V)^L_{i+1/2}$ and $F_R = (\rho V)^R_{i+1/2}$ in Fig. 3. The interpretation of the left and right state fluxes is as follows. The left state flux F_L is the contribution to the total flux from mass moving from cell i into cell $i + 1$; the right state flux F_R is the contribution from mass moving from cell $i + 1$ into cell i. Using the distribution function method, these fluxes are defined as

$$F_L = \rho^L \int_0^\infty dv\, v\, f^L(v) \tag{28}$$

and

$$F_R = \rho^R \int_{-\infty}^0 dv\, v\, f^L(v) \tag{29}$$

so that the total flux at a cell interface is

$$F = F_R + F_L \ . \tag{30}$$

In (28), only particles moving in the positive direction contribute to the flux F_L; hence, the integration limits are from 0 to ∞. Conversely, in (29), only particles moving in the negative direction contribute to the flux F_R and therefore the integration limits are from $-\infty$ to 0. Note that the fluxes are defined with variables in the left state L and the right state R. The left and right state values are determined using a high order interpolation scheme and the partial donor cell method. The high order interpolation scheme defines the values at the cell face, and the partial donor cell method is then used to define the left and right states. We now elaborate on these points.

High Order Interpolation Scheme. To calculate the fluxes at the cell faces, we need the values of the primitive variables density, velocity, and temperature (obtained from the pressure), as well as the magnetic field. The values of these variables are found by using a high order interpolation scheme. However, the high order interpolation is performed on the conserved variables mass, momentum, energy, and magnetic flux, and then converted to the appropriate primitive variables.

The scheme is described as follows [43]. We calculate the conserved variable $g_c = \int g_p\, d^3x$ where the subscripts c and p refer to conserved and primitive (e.g.,

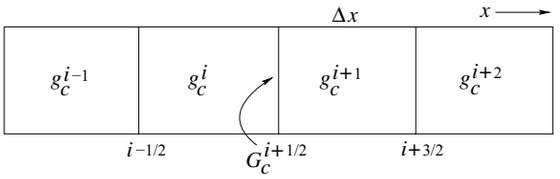


Fig. 4. One dimensional grid for the high order interpolation scheme.

total mass and density). We need to calculate the high order value G_c at the cell face. A one dimensional representation of this is shown in Fig. 4. We now define the function $\mathcal{G}(x)$ as

$$\mathcal{G}_c(x) = \int_{-\infty}^{x} G_c(s)\ ds\ . \tag{31}$$

From (31) it follows that the high order value is

$$G_c(x) = \frac{\partial \mathcal{G}_c}{\partial x}\ . \tag{32}$$

We apply (32) to find G_c to second order at the interface $i + 1/2$ and find that

$$G_{c(2)}^{i+1/2} = \left(\frac{\partial \mathcal{G}_c}{\partial x}\right)_{i+1/2} = \frac{\mathcal{G}_c^{i+3/2} - \mathcal{G}_c^{i-1/2}}{2\Delta x}\ . \tag{33}$$

Now, we note that the meaning of the last term in (33) is

$$\begin{aligned} \frac{\mathcal{G}_c^{i+3/2} - \mathcal{G}_c^{i-1/2}}{2\Delta x} &= \frac{1}{2\Delta x}\,[\text{integral of cells } i \text{ and } i+1] \\ &= \frac{1}{2\Delta x}(g_c^{i+1} + g_c^{i})\Delta x \end{aligned} \tag{34}$$

so that

$$G_{c(2)}^{i+1/2} = \frac{1}{2}(g_c^{i+1} + g_c^{i})\ . \tag{35}$$

This technique can be applied to obtain higher order interpolated values of G_c. For example, to fourth order

$$G_{c(4)}^{i+1/2} = \frac{1}{12\Delta x}\left(\mathcal{G}_c^{i-2} - 8\mathcal{G}_c^{i-1} + 8\mathcal{G}_c^{i+1} - \mathcal{G}_c^{i+2}\right) \tag{36}$$

and it can be shown that

$$G_{c(4)}^{i+1/2} = \frac{7}{12}(g_c^{i+1} + g_c^{i}) - \frac{1}{12}(g_c^{i+2} + g_c^{i-1})\ . \tag{37}$$

Finally, to obtain the high order primitive variable at the cell interface we also need to interpolate the cell volumes. For example, the primitive variable $g_{p(4)}$ is given by

$$g_{p(4)}^{i+1/2} = \frac{G_{c(4)}^{i+1/2}}{\mathcal{V}_{c(4)}^{i+1/2}} \tag{38}$$

where $\mathcal{V}_{c(4)}$ is the fourth order interpolated cell volume at $i + 1/2$.

Partial Donor Cell Method. A problem that arises in using high order interpolation schemes is that spurious overshoots and undershoots occur at sharp discontinuities (e.g., shock waves). To avoid this problem, flux limiters are used. The basic idea is that the code monitors sharp discontinuities; if a sharp discontinuity is found, then a low order interpolation scheme is used. In essence, a more diffusive scheme is used in regions of strong gradients. There are a number of flux limiting schemes available (e.g., flux corrected transport, partial donor cell method). The NRL code VooDoo uses the partial donor cell method developed in [44] which we now describe.

Consider a density structure being advected at a constant velocity V in one dimension as shown in Fig. 5. Under constant advection, the maximum density allowed is $n = n_p^{i-1}$. Since the velocity V is positive, the density structure is moving left to right and we will be determining the left state L of the density at the cell interface $i+1/2$. We retain the subscript p on the density n to emphasize that the flux limiter operates on the primitive variable not the conserved variable. In Fig. 5, n_p^i is the density in cell i, n_p^{i-1} is the density in cell $i-1$, $n_{pHO}^{i+1/2}$ is the high order interpolated density at cell interface $i+1/2$, and $n_{PDM}^{i+1/2}$ is the partial donor cell value of the density which will now be determined.

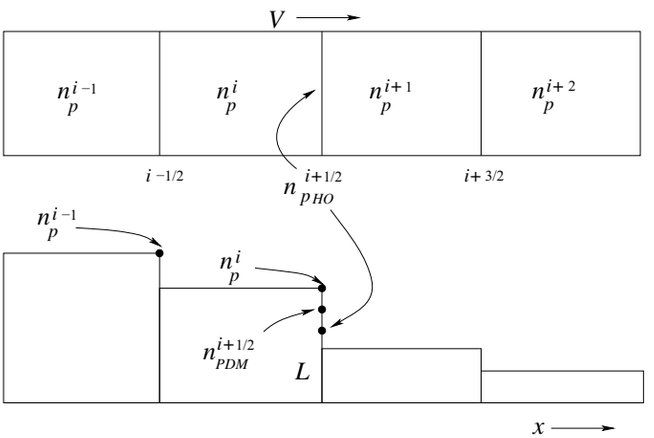


Fig. 5. Schematic of using the partial donor cell to determine the left state L of the density n.

The density is advanced one time step Δt so that the plasma moves a distance $V\Delta t$. The amount of plasma entering cell i from cell $i-1$ is $n_p^{i-1}V\Delta t$; the amount of plasma leaving cell i and going into cell $i+1$ is $n_{PDM}^{i+1/2}V\Delta t$ where we are assuming the density in the left state L is the partial donor cell value. These are represented by the lightly shaded areas in Fig. 6. Thus, the total increase in plasma in cell i after one time step is $(n_p^{i-1} - n_{PDM}^{i+1/2}V\Delta t)$. However, the maximum plasma increase allowed in cell i is $(n_p^{i-1} - n_p^i)\Delta x$ which is denoted by the darker area in Fig. 6. If more plasma enters cell i there may be an overshoot

in the density which is unphysical. The partial donor cell value is defined by balancing these two quantities, i.e.,

$$(n_p^{i-1} - n_{PDM}^{i+1/2})V\Delta t = (n_p^{i-1} - n_p^i)\Delta x \ . \tag{39}$$

From (39) we find that

$$n_{PDM}^{i+1/2} = \frac{1}{\alpha}\left[n_p^i - (\alpha - 1)n_p^{i-1}\right] \tag{40}$$

where $\alpha = V\Delta t/\Delta x$ is referred to as the PDM parameter and is in the range $0 < \alpha \leq 1$. Thus, $n_{PDM}^{i+1/2}$ is the minimum value of the left state density allowed. If the left state density is less than this value, then there is the possibility of a spurious density overshoot in cell i. Note that if $\alpha = 1$ then $n_{PDM}^{i+1/2} = n_p^i$; this is the full donor cell method and it is very diffusive numerically. By choosing $\alpha < 1$ it becomes the partial donor cell method and numerical diffusion is reduced.

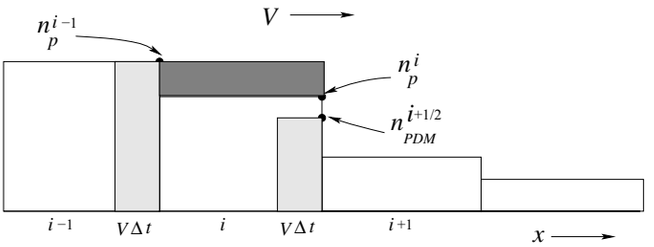


Fig. 6. Schematic showing how the partial donor cell value n_{PDM} is obtained.

Referring to Fig. 5, we see that there are three values of the density to choose from for the left state (L) density: n_p^i, $n_{PDM}^{i+1/2}$, and $n_{pHO}^{i+1/2}$. The middle value is chosen for the left state. The rationale is as follows. In general, one would want to use the high order interpolated value because it provides the best estimate of the left state. If $n_{PDM}^{i+1/2} < n_{pHO}^{i+1/2} < n_p^i$ then this would be the right choice. One would not expect a spurious overshoot in cell i because more plasma is being removed than the minimum defined by the partial donor cell value. However, for the situation shown in Fig. 5, one would choose the partial donor cell value for the left state density because $n_{pHO}^{i+1/2} < n_{PDM}^{i+1/2}$. Hence, the flux is 'limited' to prevent spurious overshoots.

To determine the right state R value of the density at cell interface i one reverses the sign of the velocity V and follows the procedure outlined above. The partial donor cell value for the right state of the density is

$$n_{PDM}^{i+1/2} = \frac{1}{\alpha}\left[n_p^i - (\alpha - 1)n_p^{i+1}\right] \ . \tag{41}$$

3.5 Distribution Function Method

The code VooDoo uses a distribution function method to calculate the fluxes of mass, momentum, and energy at cell interfaces. The question naturally arises, what is $f(v)$? This method is described in detail in [42].

We define the following transport equation

$$\frac{\partial \rho \langle \chi \rangle}{\partial t} + \nabla \cdot \rho \langle \chi \mathbf{v} \rangle - \rho \left\langle \mathbf{a} \cdot \frac{\partial \chi}{\partial \mathbf{v}} \right\rangle = 0 \tag{42}$$

where

$$\langle \mathcal{F} \rangle = \int d\mathbf{v}\, f(\mathbf{v})\, \mathcal{F}$$

$$\mathbf{a} = \frac{\nabla \mathbf{M}}{8\pi\rho}; \quad \mathbf{M} = B_x^2\, e_x + B_y^2\, e_y + B_z^2\, e_z \ .$$

The continuity equation is obtained from (42) by setting $\chi = m$, the momentum equation by setting $\chi = m\mathbf{v}$, and the energy equation by setting $\chi = mv^2/2$. The question to be answered is the following. Can we find a distribution $f(\mathbf{v})$ such that, when substituted into the transport equation (42), the 3D MHD equations (3) - (5) are recovered?

The following distribution function satisfies this requirement.

$$f = f_1 + f_2 \tag{43}$$

where

$$f_1 = \frac{\exp(-u_1^2)}{\left(\pi v_{1x}^2\right)^{1/2}} \frac{\exp(-v_1^2)}{\left(\pi v_{1y}^2\right)^{1/2}} \frac{\exp(-w_1^2)}{\left(\pi v_{1z}^2\right)^{1/2}} \tag{44}$$

$$f_2 = -\left(u_2 v_2 + u_2 w_2 + v_2 w_2\right) \frac{\exp(-u_2^2)}{\left(\pi v_{2x}^2\right)^{1/2}} \frac{\exp(-v_2^2)}{\left(\pi v_{2y}^2\right)^{1/2}} \frac{\exp(-w_2^2)}{\left(\pi v_{2z}^2\right)^{1/2}} \tag{45}$$

and $u_\alpha = (v_x - V_x)/v_{\alpha x}$, $v_\alpha = (v_y - V_y)/v_{\alpha y}$, $w_\alpha = (v_z - V_z)/v_{\alpha z}$, $v_{1x}^2 = 2C_s^2/\gamma + V_{Ay}^2 + V_{Az}^2$, $v_{1y}^2 = 2C_s^2/\gamma + V_{Ax}^2 + V_{Az}^2$, $v_{1z}^2 = 2C_s^2/\gamma + V_{Ax}^2 + V_{Ay}^2$, $v_{2x}^2 = 4V_{Ax}^2$, $v_{2y}^2 = 4V_{Ay}^2$, $v_{2z}^2 = 4V_{Az}^2$, and $C_s^2 = \gamma P/\rho$. In (43), f_1 defines the diagonal terms in the MHD equations and f_2 defines the off-diagonal terms.

The point of this method is that the distribution function $f(v)$ contains the hydrodynamic force (pressure) and all of the electromagnetic forces except those associated with $\mathbf{a}$. Thus, by using $f(v)$ to calculate mass, momentum, and energy fluxes, most of the electromagnetic contribution is calculated directly, i.e., one does not have to solve the hydrodynamic and magnetic force components separately or differently.

3.6 Magnetic Field Evolution

The evolution of the magnetic field is governed by

$$\frac{\partial \mathbf{B}}{\partial t} = -c\nabla \times \mathbf{E} = -c\nabla \times [\mathbf{E}_c + \mathbf{E}_H] \tag{46}$$

where we define $\mathbf{E}_c = -\mathbf{V} \times \mathbf{B}/c$ and $\mathbf{E}_H = -\mathbf{V}_H \times \mathbf{B}/c$ with $\mathbf{V}_H = -\mathbf{J}/ne$. The electric field $\mathbf{E}_c$ is the convective electric field and $\mathbf{E}_H$ is the Hall electric field. The general technique to solve (46) is as follows. The normal component of the field is integrated over a cell face

$$\int d\mathbf{A}\cdot\frac{\partial \mathbf{B}}{\partial t} = -c\int d\mathbf{A}\cdot \nabla \times \mathbf{E} = -c\int \mathbf{E}\cdot\ d\mathbf{l} \tag{47}$$

where $\int d\mathbf{l}$ denotes a line integral around the cell face (see Fig. 1). The advantage of using this technique to calculate the magnetic field is that is satisfies $\nabla\cdot\mathbf{B} = 0$ to machine accuracy.

As a specific example, the magnetic flux in the x direction at a cell interface is evolved using

$$\frac{\partial \Phi_x}{\partial t} = -cE_y\, dy - cE_z\, dz \tag{48}$$

where $\Phi_x = B_x\, dydz$. The updated magnetic flux in the x direction is

$$\Phi_x^{t+\Delta t} = \Phi_x^t + c\Delta t\,(E_y\, dy + E_z\, dz)^{t+\Delta t/2}\ . \tag{49}$$

The convective and Hall electric fields are structurally very different and require different numerical techniques. We describe each field calculation separately.

Convective Electric Field. The convective electric field in VooDoo is calculated using the distribution function method. This is most easily seen by considering a single component of the convective electric field

$$E_{cz} = -\frac{1}{c}(V_xB_y - V_yB_x) \tag{50}$$

which is written as

$$E_{cz} = -\frac{1}{c}\left(\int f(\mathbf{v})v_xB_y\, d^3v - \int f(\mathbf{v})v_yB_x\, d^3v\right)\ . \tag{51}$$

Thus, the calculation of $\mathbf{E}_c$ is based on the fact that $\mathbf{V} = \int f(\mathbf{v})\mathbf{v}\, d^3v$. We note that the actual implementation of this technique is fairly complicated; it is described in [42].

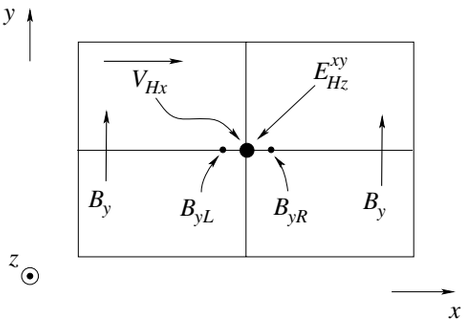


Fig. 7. Schematic of the upwinding scheme used to calculate one component of the Hall electric field.

Hall Electric Field. The Hall electric field is

$$\mathbf{E}_H = -\frac{1}{c}\mathbf{V}_H \times \mathbf{B} = \frac{1}{nec}\mathbf{J} \times \mathbf{B} \ . \tag{52}$$

Since the Hall electric field is solely a function of the magnetic field, the distribution function method cannot be used to solve for it. We use an upwinding scheme to solve for the Hall electric field. We will explain this method by using a simple example.

We consider a single component of the Hall electric field

$$E_{Hz} = -\frac{1}{c}(V_{Hx}B_y - V_{Hy}B_x) \tag{53}$$

which we break up into two parts

$$E^{xy}_{Hz} = -V_{Hx}B_y \quad \text{and} \quad E^{yx}_{Hz} = V_{Hy}B_x \ . \tag{54}$$

By analogy with the convective electric field, we interpret B_y being convected at the Hall velocity V_{Hx} and B_x being convected at the Hall velocity V_{Hy} in (54).

We now focus on the component E^{xy}_{Hz} which is shown in Fig. 7. In this figure E^{xy}_{Hz} and V_{Hx} are defined along a cell edge (as shown in Fig. 1) and $V_{Hx} > 0$. Also shown are the face-centered magnetic field B_y and the left (B_{yL}) and right (B_{yR}) states of B_y. The left and right states of B_y are determined using the high order interpolation scheme and partial donor cell method previously described. The upwinding scheme to determine E^{xy}_{Hz} is

$$E^{xy}_{Hz} = \begin{cases} V_{Hx}B_{yL} & \text{for } V_{Hx} > 0 \\ V_{Hx}B_{yR} & \text{for } V_{Hx} < 0 \end{cases} \tag{55}$$

The interpretation of (55) is straightforward. If $V_{Hx} > 0$, as shown in Fig. 7, then the Hall term is convecting the magnetic field from left to right and the left state magnetic field is used, i.e., the upwinded state. Conversely, if $V_{Hx} < 0$, then the right state magnetic field is used.

3.7 Courant Condition

The time step in an explicit code is controlled by a Courant condition

$$\Delta t < \frac{\Delta x}{V_{max}} \tag{56}$$

where V_{max} is the maximum fluid or wave velocity in the system. Usually the Courant condition is set by wave speeds. The following velocities need to be considered in a Hall MHD code:

- Fluid flow: V
- Magnetosonic wave: $V_m = \sqrt{V_A^2 + C_s^2}$
- Hall velocity: $V_H = -J/ne$
- Hall drift wave: $V_{hdw} = V_A^2/L_n\Omega_i$
- Whistler wave: $V_w = kV_A^2/\Omega_i$

where $V_A = B/(4\pi n_i m_i)^{1/2}$ is the Alfvén velocity, $C_s = (2T/m_i)^{1/2}$ is the sound speed, $L_n = (\partial ln\, n/\partial x)^{-1}$ is the density gradient scale length, $\Omega_i = eB/m_i c$ is the ion cyclotron frequency, and k is the wavenumber. We set $k = 2\pi/\Delta x$ to evaluate V_w numerically. Thus, we use the following velocity in the Courant condition

$$V_{max} = \max(V + V_m + V_H + V_{hdw} + V_w)\ . \tag{57}$$

In general, the Hall drift and whistler wave phase velocities determine the time step. For phenomena in which only the plasma dynamics in the plane transverse to the magnetic field is important, the Hall drift wave controls the time step; examples are plasma opening switches and sub-Alfvénic plasma expansions. On other hand, for phenomena that require plasma dynamics along the magnetic field, the whistler wave phase velocity controls the time step; a prominent example is magnetic field line reconnection.

The whistler wave is particularly troublesome in modeling Hall MHD processes for two reasons. First, $V_w >> V_A$ so that the time step is significantly smaller than that used in ideal MHD simulations. Second, and most important, the whistler phase velocity is proportional to k (or $1/\Delta x$ numerically) so that the shortest waves in the system have the highest phase velocity. This can generate spurious, short wavelength noise in Hall MHD simulations. This noise can be suppressed by adding an artificial hyper-resistivity to the magnetic induction equation or by numerically smoothing the Hall electric field. The latter is done in VooDoo using a 3-point smoothing algorithm [45]. In addition, there is also a severe penalty in the time step when the resolution is increased. The time step is proportional to $(\Delta x)^2$ so that time step decreases by an additional factor of 2 when the resolution is doubled.

3.8 Sub-cycling the Hall Physics

Hall MHD simulations can be very computationally intensive because of the small time step required. One technique to overcome the small time step problem

is to solve the magnetic field equation for the Hall term implicitly. Although this will allow for larger time steps and probably eliminate spurious, short-wavelength noise, it can suppress important physical processes. An alternative technique, used in VooDoo, is to sub-cycle the Hall term.

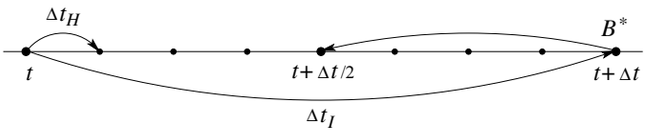


Fig. 8. Schematic showing the sub-cycling of the Hall term.

One subcycling time advance scheme is shown in Fig. 8. The ideal time step Δt_I is defined by the Courant condition using $V_{max} = \max(V + V_m)$ while the Hall time step Δt_H is defined using (57). The basic idea is that the plasma does not move on the Hall time scale (i.e., $V = 0$) so we do not need to solve the full set of equations (3) - (6) on the Hall time scale. We only solve (52) on the Hall time scale. We advance the magnetic field using (52) on the Hall time step scale Δt_H from t to $t + \Delta t$ and obtain a provisional value of B^*. We use B^* at the half-time step and do a full ideal MHD advance from t to $t + \Delta t$. The ideal MHD equations (3) - (5) are solved along with the convective magnetic induction equation

$$\frac{\partial \mathbf{B}}{\partial t} = -c\nabla \times \mathbf{E}_c = \nabla \times \mathbf{V} \times \mathbf{B} \ . \tag{58}$$

This method is used in VooDoo and it substantially reduces the computational time required for a 3D Hall MHD simulation (by about an order of magnitude).

4 Applications

We now present several applications of Hall MHD to space and laboratory plasmas. We first look at the propagation of whistler and Hall drift waves, and then examine plasma opening switches and magnetic reconnection.

4.1 Linear Hall Waves

In developing a Hall MHD code it is important that the two fundamental linear wave modes are properly described. We present numerical results of the dispersion relation of whistler and Hall drift waves using VooDoo.

Whistler Waves. The linear wave relation for whistler waves is given by (15). We compare the numerical results of a simulation with this dispersion relation. The simulation parameters are the following. The ambient magnetic field is in the z-direction $\mathbf{B} = B_0\,\hat{\mathbf{e}}_z$ with $B_0 = 1000$ G, the density n is homogeneous with n_0

$= 10^{12}$ cm^{-3}, $\beta = 10^{-4}$, the plasma is assumed to be isothermal (i.e., $T =$ cnst), the system size is $L_z = 20$ cm, and we use 120 mesh points in the z-direction. The system is perturbed with $\delta B_x = \delta B \sin(2\pi mz/L_z)$ and $\delta B_y = \delta B \cos(2\pi mz/L_z)$ where m is the mode number and we take $\delta B = 10$ G. In Fig. 9 we plot the wave frequency as a function of mode number for the analytical and numerical results. The quadratic nature of the wave mode is evident. At low mode number the the two results agree extremely well. However, at high mode number the numerical frequency becomes increasingly smaller than the analytical frequency because of grid dispersion, i.e., the wavelength of the mode approaches the grid scale.

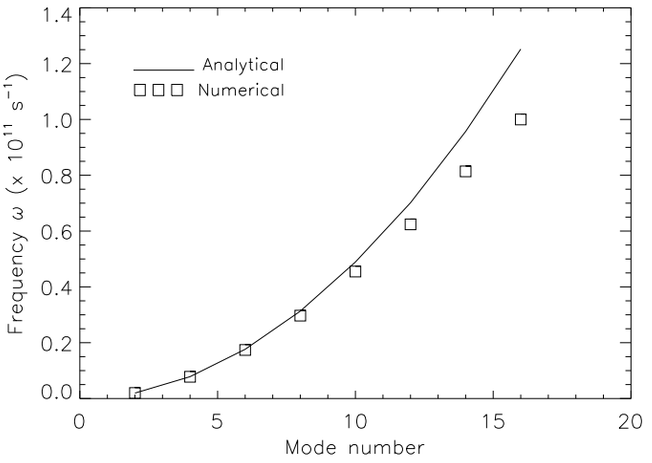


Fig. 9. Analytical and numerical comparison of the whistler wave frequency.

Hall Drift Waves. A simple linear wave relation for Hall drift waves is given by (18) that is valid in the limit $L_n << c/\omega_{pi}$. However, the more general dispersion equation is given by [38]

$$\omega^2 - k_y^2 V_A^2 - \omega k_y V_A^2/\Omega_{ci} L_n = 0 \tag{59}$$

for $\beta << 1$. The solution to this equation is

$$\omega = \frac{k_y V_A^2}{2L_n \Omega_{ci}} \pm \frac{1}{2}\left[\frac{k_y^2 V_A^4}{L_n^2 \Omega_{ci}^2} + 4k_y^2 V_A^2\right]^{1/2} \tag{60}$$

which illustrates the coupling between the Alfvén wave and the Hall drift wave. In the limit $L_n >> V_A/\Omega_{ci}$ (or $L_n >> c/\omega_{pi}$), (60) reduces to $\omega = k_y V_A$; in the opposite limit it reduces to (18). We now present simulation results for this wave mode.

The simulation is set up as follows. We consider a 2D grid in the $x - y$ plane; the length scales of the system are $L_x = 30$ cm and $L_y = 20$ cm and the

mesh size is $(x, y) = (100,120)$. The ambient magnetic field is in the z-direction ($\mathbf{B} = B_0\,\hat{\mathbf{e}}_z$) with $B_0 = 1000$ G, the density is inhomogeneous with a profile

$$n(x) = \frac{n_0}{2}\,[(1 + A) + (1 - A)\tanh((x - x_0)/\Delta x)] \tag{61}$$

where A is referred to as the Atwood number, $\beta = 10^{-4}$, and the plasma is assumed to be isothermal (i.e., $T = $ cnst). The magnetic field is perturbed with $\delta B_z = \delta B \cos(2\pi m y/L_y)$ where m is the mode number and we take $\delta B = 10$ G. We also use $A = 20$, $x_0 = 10$ cm, and $\Delta x = 3$ cm. Finally, we maintain the density profile (61) throughout the simulation to maintain the propagation of the Hall drift wave. For this plasma configuration the Hall drift wave propagates in the $-y$-direction.

In the left panel of Fig. 10 we plot the normalized phase velocity of the Hall drift wave and the normalized density as a function of x. The Hall drift wave dominates the Alfvén wave in the region of the density gradient. The maximum phase velocity of the Hall drift wave occurs at $x \simeq 15$ cm. In the right panel of Fig. 11 we show a contour plot of the perturbed magnetic field $B(x, y) - B_0$ at time $t = 10^{-8}$ s. Note that the contours are distorted in the $-y$-direction in the region of the density gradient, and that the maximum distortion occurs at $x \simeq 15$ cm. This is consistent with the linear theory of the Hall drift wave.

In Fig. 11 we plot the maximum wave frequency of the Hall drift wave (which is at $x \simeq = 15$ cm) as a function of mode number. The comparison between the analytical result and the numerical results is good. Note that the wave dispersion is linear and not quadratic like the whistler wave.

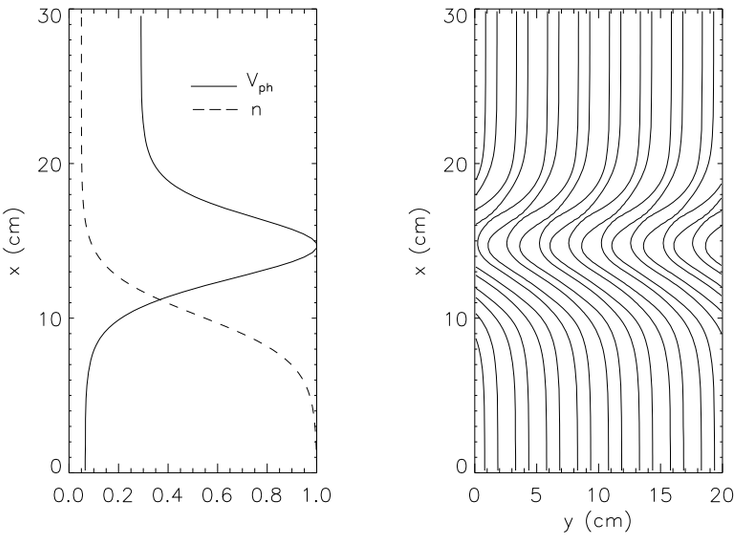


Fig. 10. Plot of the normalized phase velocity and density (left panel) and contour plot of the perturbed magnetic field (right panel) for the Hall drift wave.

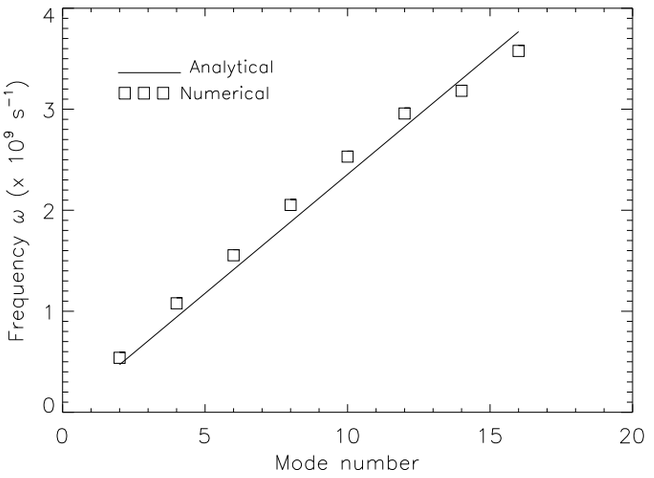


Fig. 11. Analytical and numerical comparison of the Hall drift wave frequency at $x = 15$ cm.

4.2 Plasma Opening Switch

The plasma opening switch is a laboratory device designed to provide the appropriate power conditioning for inductive, pulsed, high-power generators [8,9,10,13,14,17]. A schematic of an opening switch is shown in Fig. 12. The basic concept of the switch is that a capacitive discharge sets up a current system through the plasma in the switch. The electrostatic energy stored in the capacitor is converted to magnetic inductive energy. The discharge current propagates down the plasma channel. When it reaches the end of the plasma the switch 'opens' and it rapidly transfers the stored magnetic energy to the load. The energy released is in the range 10 – 1000 KJ over a time scale 10's ns to a few μs so that the power produced is 10^{12} – 10^{14} W. Depending on the plasma parameters, Hall physics can control the dynamics of current propagation down the channel. The key plasma parameter that controls the dynamics is the plasma density in the switch. At high densities $n > 10^{16}$ cm^{-3} the plasma is controlled by ideal MHD (i.e., Alfvén time scales), while at low densities $n < 10^{13}$ cm^{-3} the plasma is dominated by Hall physics. At intermediate densities, both ideal MHD and Hall physics are important.

We present results for two plasma opening switch simulations. The simulation is set up as follows. We consider a 2D grid in the $x-y$ plane; the length scales of the system are $L_x = 8$ cm and $L_y = 8$ cm and the mesh size is $(x, y) = (40{,}50)$; the mesh is nonuniform in the y-direction. The density profile is

$$n(y) = \frac{n_0}{Z}\left[(1 + \alpha \tanh(y/\Delta y)\right] \tag{62}$$

where Z is the charge state. We choose $\alpha = 9$ and $\Delta y = 0.5$ for $y < 0$, and $\alpha = 3$ and $\Delta y = 1.0$ for $y > 0$. A time-dependent magnetic field in the z-direction

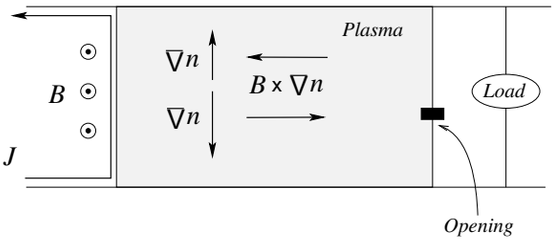


Fig. 12. Schematic of a plasma opening switch.

($\mathbf{B} = B_0 \tanh(t/\tau_r)\,\hat{\mathbf{e}}_z$) is imposed at the $x = 0$ boundary with $B_0 = 1.2 \times 10^4$ G and $\tau = 10^{-6}$ s; this models the current rise time associated with the capacitive discharge. Finally, the plasma is assumed to be isothermal (i.e., T = cnst) and consists of doubly charged carbon ions ($Z = 2$ and $m_i = 12 m_p$). We also assume $\beta = 10^{-4}$,

The results are shown in Fig. 13. In the left panel we plot the normalized phase velocity V_{ph} of the Hall drift wave (solid line) and normalized density n (dashed line) for $n_0 = 10^{12}$ cm^{-3}. The magnitude of V_{ph} is a function of space and time because of the time dependent magnetic field. The interesting feature of this system is that $V_{ph} > 0$ for $y < 0$ and $V_{ph} < 0$ for $y > 0$ because the Hall drift propagates in the $\mathbf{B} \times \nabla n$ direction. Hence, in the Hall regime, the magnetic field will penetrate the plasma for $y < 0$ and be inhibited for $y > 0$. In the middle panel we show a contour plot of the magnetic field for $n_0 = 10^{12}$ cm^{-3} at time t = 20 ns. The magnetic field has penetrated the system extremely rapidly in the region where the Hall drift hall velocity is positive. In contrast, the time scale for penetration for ideal MHD is roughly two orders of magnitude longer ($\simeq 2\mu s$). For $B_0 \simeq 200$ G we find that $V_A \simeq 6 \times 10^7$ cm/s and $V_{ph} \simeq 8 \times 10^9$ cm/s. In the right panel we show results for $n_0 = 10^{14}$ cm^{-3}. The Hall penetration time of the magnetic field is an order of magnitude longer the previous case because of the increased density. Also, the field penetration 'tongue' is broader than the previous case. This is attributed to ideal MHD effects, i.e., the magnetic field is able to 'push' it's way into the plasma.

4.3 Magnetic Reconnection

In recent years it has become evident that Hall physics plays a critical role in magnetic reconnection processes [18,19,20,21,22,23,24,25,26,27,28,29,30,31]. This is especially true at the magnetopause and in the magnetotail where reconnection processes can dominate the dynamics of the system. The Hall term decouples the ion and electron motion on scale lengths less than an ion inertial length c/ω_{pi}. Drake and co-workers [20,21,23,28] made the observation that the ion outflow channel in the reconnection process is therefore determined by c/ω_{pi} rather than the electron scale length c/ω_{pe}. This is significant because it allows reconnection to proceed at a rapid rate. We note that nominal values of the ion inertial length at the earth's magnetopause and in the magnetotail are

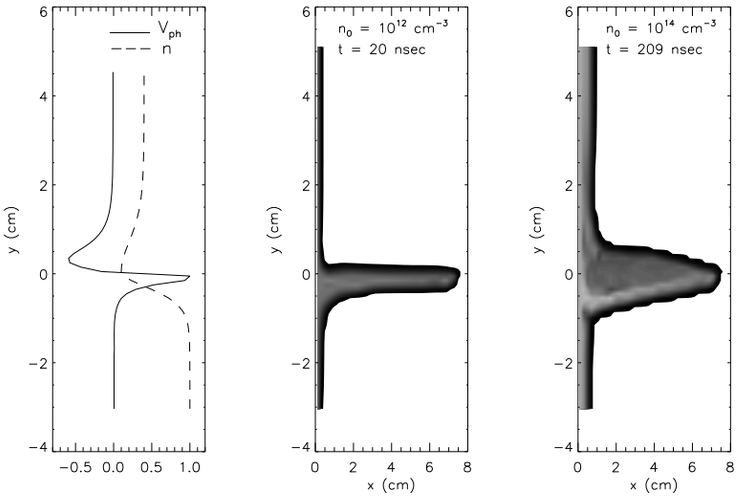


Fig. 13. Simulation of a plasma opening switch.

$c/\omega_{pi} \sim 10$'s – 100's km. This is much smaller than grid sizes used in global MHD models of the earth's magnetosphere and highlights a computational difficulty in incorporating Hall physics into a global model of the magnetosphere.

We present a comparison of 2D Hall and ideal MHD simulations using the parameters defined for the GEM Reconnection Challenge problem [24]. The magnetic field is in the x-y plane. Periodic boundary conditions are used in the x-direction and zero-gradient ($\partial/\partial y = 0$) boundary conditions in the y-direction. The equilibrium magnetic field is given by $B_x(y) = B_0 \tanh(y/w)$ with $w = 0.5c/\omega_{pi}$ and c/ω_{pi} is the ion inertial length defined using $n = n_0$. The density floor n_b is 1/5 of the maximum density n_0. The temperature T is chosen to provide pressure balance; the parameters n_0 and B_0 are adjusted so that $\beta_0 = 8\pi n_0 T/B_0^2 = 1$. The simulations are initialized with a large magnetic island with a flux perturbation of the form $\psi = -\psi_0 \cos(2\pi x/L_x)\cos(\pi y/L_y)$ where L_x and L_y are the dimensions of the simulation box. The size of the simulation box is $L_x = 25.6c/\omega_{pi}$ and $L_y = 12.3c/\omega_{pi}$, and we take $\psi_0 = 0.1$. The time scale is normalized to $\tau_A = (c/\omega_{pi})/V_{A0}$ based on $n = n_0$ and $B = B_0$. The mesh size of the simulations are 100×50 and uses a stretched mesh in the y-direction. There are roughly 25 grid points within the current layer.

A significant difference between Hall and ideal MHD reconnection dynamics is the development of an 'out-of-plane' magnetic field B_z. This component of the wave field is associated with the generation of a whistler wave. This is shown in Fig. 14 at time $t \simeq 21\tau_A$; this field does not develop in ideal MHD reconnection. The origin of this field and its structure follows from (52). The z-component of this equation is nonzero and is given by

$$\frac{\partial B_z}{\partial t} = -\frac{\partial}{\partial x}\frac{J_z B_x}{ne} - \frac{\partial}{\partial y}\frac{J_z B_y}{ne} \tag{63}$$

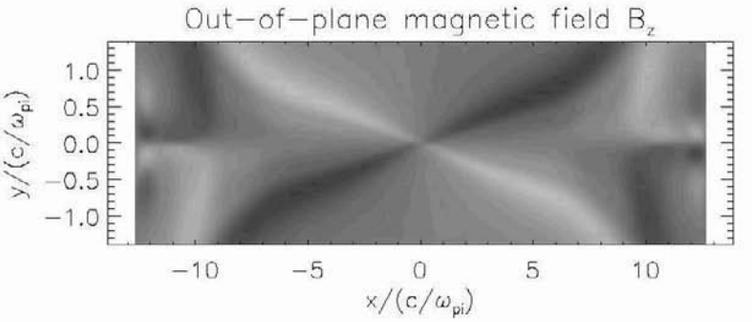


Fig. 14. Out of plane magnetic field that develops during magnetic reconnection for Hall MHD.

where we assume that reconnection has commenced and that $B_z \simeq 0$, $J_x \simeq 0$, and $J_y \simeq 0$. It is found that the first term reverses sign at $x = 0$ and that the second term reverses sign at $y = 0$. Hence, because of these asymmetries the 'out-of-plane' magnetic field B_z develops a quadrupole structure as shown in Fig. 14. The generation of B_z subsequently causes an 'out-of-plane' flow V_z because of the $\mathbf{J} \times \mathbf{B}$ force that now develops in the z-direction (e.g., $J_x \propto \partial B_z/\partial y$ and $J_y \propto \partial B_z/\partial x$). We note that for ideal MHD, one replaces J_z/ne with V_z in (63); if $V_z = 0$ initially then it will remain zero in ideal MHD.

The development of the 'out-of-plane' current J_z is also dramatically different as shown in Fig. 15. In Fig. 15 we show contour plots of the current J_z for ideal and Hall MHD at time $t \simeq 21\tau_A$. The ideal MHD case forms a thin current layer the length of the system, while the Hall MHD case forms a distinctive X-point structure current system. Again, this is associated with the development of the whistler wave field. We point out that subsequently the ideal MHD current

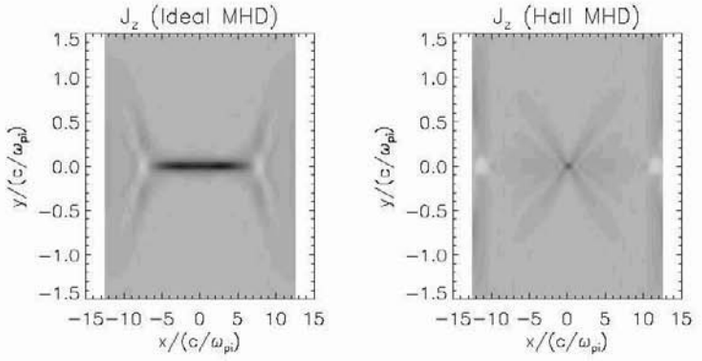


Fig. 15. Out of plane current J_z that develops during magnetic reconnection for ideal and Hall MHD.

layer becomes unstable to secondary reconnection processes, while the Hall MHD current layer remains stable.

Perhaps most important, the Hall MHD magnetic reconnection rate is substantially faster the ideal MHD reconnection rate. In Fig. 16 we plot the reconnected flux ($\Phi = \int B_y\, dx$ where the line integration is between the X point and the O point) as a function of time for the Hall and ideal MHD cases. The Hall MHD case reconnects substantially faster than the ideal MHD case. This is consistent with other simulation studies (e.g., see [24]).

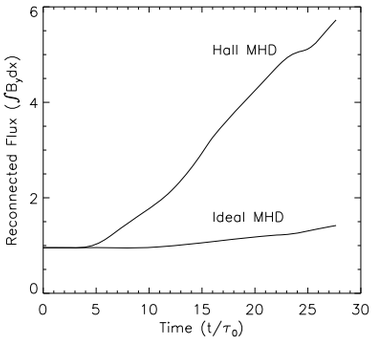


Fig. 16. Reconnected magnetic flux as a function of time.

A number of researchers are attacking the reconnection problem in 3D using particle [32,33], hybrid [34], and fluid codes [35]. We now present some results of a 3D simulation study of Hall reconnection physics using VooDoo [36].

The simulation conditions for the main study are as follows. Periodic boundary conditions are used in the x direction and zero-gradient boundary conditions in the y and z directions ($\partial/\partial y = 0$ and $\partial/\partial z = 0$). The spatial scales are normalized to the the ion inertial length (c/ω_{pi0}), the time scale to the ion gyrofrequency (Ω_{i0}), and the velocity to the Alfvén velocity (V_{A0}) using $n = n_0$ and $B = B_0$. The equilibrium magnetic field is given by $B_x(y) = B_0 \tanh(y/y_L)$ with $y_L = 0.5$. The temperature is defined to be $C_s = V_{A0}$ where $C_s = (2T/m_i)^{1/2}$. The density profile is determined by balancing the plasma and magnetic field pressures. The maximum density is $n = n_0$ at $y = 0$ and is $n = 0.2n_0$ for $|y| > y_L$. The size of the simulation box is $L_x = 26$, $L_y = 12.5$, and $L_z = 104$. The simulations are initialized with a magnetic perturbation $\delta B_x = (\delta B/2)(L_x/L_y)F(x,y)$ and $\delta B_y = \delta B F(x,y)$ localized in the z direction $-26.8 < z < -20.4$ so that the width of perturbation is $\Delta \simeq 6$. We take $F(x,y) = \sin(2\pi x/L_x)\cos(\pi y/L_y)$ and $\delta B = 0.1B_0$. The mesh size is $50 \times 50 \times 50$. A nonuniform stretched mesh is used in the y-direction so that there are roughly 25 grid points within the current layer. The simulations are run to time $\tau = 28.6$.

The results of this simulation are shown in Figs. 17 – 19. In Fig. 17 we show contours of the plasma density and velocity vectors as a function of space (yz plane at $x = 0$) at times $t = 1.6, 16.4, 32.7,$ and 49.1. Here, the velocity vectors

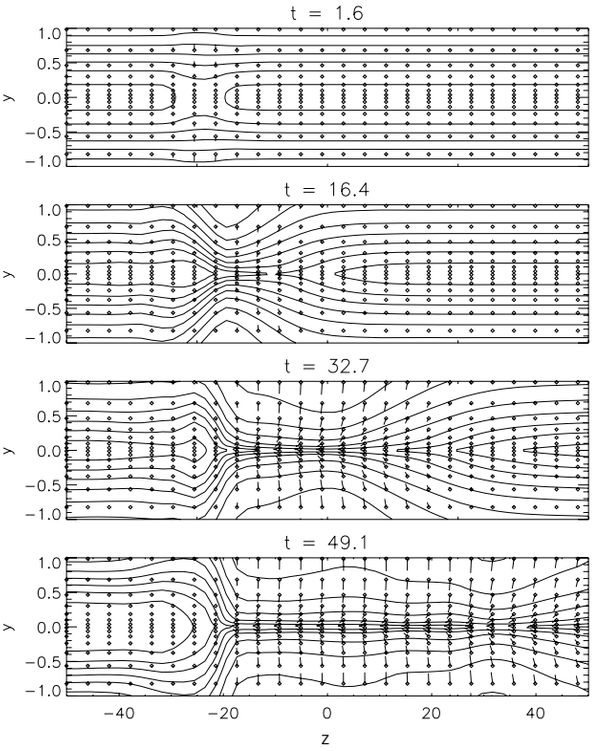


Fig. 17. Contour plots of the plasma density and velocity vectors at times t = 1.6, 16.4, 32.7, 49.1.

are 'wind flags' where the small diamond is at the base of the vector. Three observations are made. First, although the initial perturbation width is $\Delta \simeq 6$, the reconnection layer extends a distance of $\Delta \simeq 70$ at $t = 49.1$. Second, the reconnection site propagates asymmetrically; it only propagates in the $+z$ direction from it's initial position at $z \sim -22$. The disturbance does not propagate into the region $z < -30$. And third, as the reconnection wave propagates in the $+z$ direction, plasma flows are not only directed toward the neutral line but also in the direction of the current (i.e., the $-z$ direction) at the front of the wave; this is evident at time $t = 32.7$. However, after the reconnection wave has passed through the system, the resulting flow pattern away from the neutral line is similar to the 2D case, i.e., plasma flows are only directed toward the neutral line. at $x = 0$. This is shown at time $t = 49.1$ in the region $-20 < z < 30$.

In Fig. 18 we show magnetic streamlines at $z = 0$ and $\tau = 28.6$. The magnetic field lines that have just reconnected are bent (or stretched) in the $+z$ direction. After the field lines reconnect, magnetic tension 'pulls' them toward the O point, but because there is also a component of the tension in the $-z$ direction, the reconnected field lines overshoot the position $z = 0$ and become bent in the $-z$

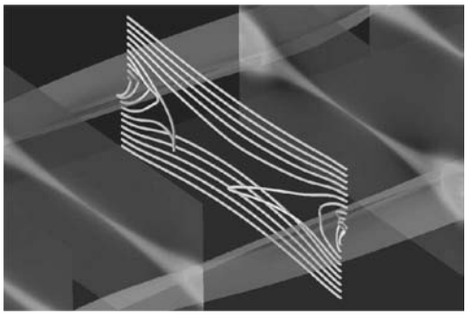


Fig. 18. Streamlines of the magnetic field at time $\tau = 28.6$ and position $z = 0$.

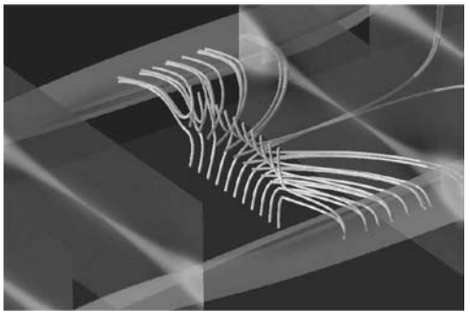


Fig. 19. Streamlines of the flow velocity at time $\tau = 28.6$ and position $z = 0$.

direction near the O point. This is consistent with the 2D simulation results shown in Fig. 14.

Figure 19 shows the consequences of this reconnection process on the plasma flow. In Fig. 19 we show the plasma flow streamlines of plasma that originated at $z = 0$ and $y = 0$ and ± 5 at several points along the x axis. The plasma initially flows toward the midplane, i.e., in the $\pm y$ direction, as shown in Fig. 17. Plasma that originated close to the X point (i.e., $x \simeq 0$) is accelerated primarily in the $-z$ direction. This is shown by the streamlines in the center of the plasma sheet that extend to $z < -20$. However, as one moves away from the X point, the plasma flow 'fans out' and there is flow in both the $\pm x$ direction and the $-z$ direction. The reconnecting field lines appear to 'whip' the plasma towards the O points as they release their tension.

The dynamics of the magnetic field in the reconnection process (i.e., asymmetric propagation of the reconnection layer and field line bending) is explained as follows. For simplicity we rewrite Faraday's law [Eq. (6)] in the electron magnetohydrodynamic limit ($\mathbf{V} = 0$) as

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times \left[\frac{c}{4\pi n e} \left(\nabla B^2/2 - (\mathbf{B} \cdot \nabla)\mathbf{B} \right) \right] . \tag{64}$$

When the magnetic field is uni-directional, i.e., no magnetic curvature, only the first term in Eq. (64) contributes to wave propagation in the Hall limit.

In this limit a density gradient is required to support the Hall drift wave; the consequences of this wave mode on the evolution of a plasma current sheet was discussed in Rudakov and Huba [41]. However, in a reconnection geometry near the neutral sheet, magnetic field line curvature dominates and the second term in Eq. (64) controls the evolution of the magnetic field in the Hall limit. The evolution of B_y is approximately described by

$$\frac{\partial B_y}{\partial t} \simeq -\frac{\partial}{\partial z}\frac{cB_y}{4\pi ne}\frac{\partial B_x}{\partial y} \simeq \frac{J_z}{ne}\frac{\partial B_y}{\partial z} \ . \tag{65}$$

We can rewrite Eq. (65) as $\partial B_y/\partial\zeta = 0$ where $\zeta = z - V_B t$ and $V_B = (c/4\pi ne)\partial B_x/\partial y = -J_z/ne$. Thus, the y component of the magnetic field propagates in the direction opposite to the current at a velocity $\sim V_B$; this 'reconnection wave' is responsible for the asymmetric propagation of the reconnection layer (and is also responsible for the rapid penetration of magnetic flux in cylindrical plasma opening switches [9]). Using (65) we estimate $V_B \simeq 1.6V_{A0}$ where $V_{A0} = B_0/(4\pi n_0 m_i)^{1/2}$ for plasma conditions near the X point in the simulation study. This is in reasonable agreement with the wave velocity obtained from the simulation $V_B \simeq 2.0V_{A0}$. The wave speed based on Eq. (65) underestimates the wave speed because ion motion is neglected.

5 Summary

We have presented a brief tutorial on Hall magnetohydrodynamic (Hall MHD) physics. The emphasis has been on a basic description of the new wave modes introduced by the Hall term (the whistler wave and the Hall drift wave), and on numerical methods to introduce the Hall term into an MHD code. In addition, an overview of the fundamental methods used in the NRL 3D Hall MHD code VooDoo is presented. Two applications of Hall MHD physics are presented. The plasma opening switch which illustrates rapid magnetic field transport associated with Hall drift wave, and magnetic field line reconnection which demonstrates that the Hall term can dramatically enhance the reconnection rate and can lead to the propagation of a 'reconnection wave' in a three-dimensional system.

Acknowledgments

I am indebted to Dr. J. Lyon who developed a 2D Hall MHD code at NRL in the mid-80's; the basic transport scheme for the Hall term used in VooDoo and described in this paper is based on his original work. I wish to thank several of my colleagues for valuable discussions over the years: Steve Zalesak, Glenn Joyce, Leonid Rudakov, Clark Mobarry, Joel Fedder, Dan Winske, and Adil Hassam. This research has been supported by the Office of Naval Research and the National Aerospace and Space Administration.

References

1. P.A. Bernhardt, R.A. Roussel-Dupré, M.B. Pongratz, G. Haerendel, A. Valenzuela, D. Gurnett, and R.R. Anderson: J. Geophys. Res. **92**, 5777 (1987).
2. A.B. Hassam and J.D. Huba: Geophys. Res. Lett. **14**, 60 (1987).
3. Ripin, B.H., E.A. McLean, C.K. Manka, C. Pawley, J.A. Stamper, T.A. Peyser, A.N. Mostovych, J. Grun, A.B. Hassam, and J.D. Huba: Phys. Rev. Lett. **59**, 2299 (1987).
4. J.D. Huba, J.G. Lyon, and A.B. Hassam: Phys. Rev. Lett. **59**, 2971 (1987).
5. G. Dimonte and L.G. Wiley: Phys. Rev. Lett. **67**, 1755 (1991).
6. J.D. Huba, P.A. Bernhardt, and J.G. Lyon: J. Geophys. Res. **97**, 11 (1992).
7. Ripin, B.H., J.D. Huba, E.A. McLean, C.K. Manka, T.A. Peyser, H.R. Burris, and J. Grun: Phys. Fluids B **5**, 3491 (1993).
8. R.A. Meger, R.J. Commisso, G. Cooperstein, and S.A. Goldstein: Appl. Phys. Lett. **42**, 943 (1983).
9. B. V. Weber, R. J. Commiso, R. A. Meger, J. M. Neri, W. F. Oliphant and P. F. Ottinger, Appl. Phys. Lett. **45**, 1043 (1984).
10. G. Cooperstein and P.F. Ottinger: Guest Editorial, IEEE Trans. Plasma Sci. **PS-15**, 629 (1987).
11. A.S. Kingsep, K.V. Chukbar, and V.V. Yan'kov: Reviews of Plasma Physics, Vol. 16, ed. B.B. Kadomtsev, Consultants Bureau, New York, 1990.
12. A. Fruchtman: Phys. Fluids B **3**, 1908 (1991).
13. R.J. Commisso, P.J. Goodrich, J.M. Grossmann, D.D. Hinshelwood, P.F. Ottinger, and B.V. Weber: Phys. Fluids B **4**, 2368 (1992).
14. D.D. Hinshelwood, B.V. Weber, J.M. Grossmann, and R.J. Commisso: Phys. Rev. Lett. **68**, 3567 (1992).
15. R.J. Mason, P.L. Auer, R.N. Sudan, B.V. Oliver, C.E. Seyler, and J.B. Greenly: Phys. Fluids B **5**, 1115 (1993).
16. J.D. Huba, J.M. Grossmann, and P.F. Ottinger: Phys. Plasmas **1**, 3444 (1994).
17. A. S. Chuvatin and B. Etlicher: Phys. Rev. Lett. **74**, 2965 (1995).
18. J.F. Drake, R.G. Kleva, and M.E. Mandt: Phys. Rev. Lett. **73**, 1251, 1994.
19. D.E. Biskamp, Schwarz, and J.F. Drake: Phys. Rev. Lett. **75**, 3850, 1995.
20. M.A. Shay, J.F. Drake, R.E. Denton, and D. Biskamp: J. Geophys. Res. **103**, 9165, 1998.
21. M.A. Shay, and J.F. Drake: Geophys. Res. Lett. **25**, 3759, 1998.
22. M. Rastätter, M. Hesse, and K. Schindler: J. Geophys. Res. **104**, 12,301, 1999.
23. M.A. Shay, J.F. Drake, B.N. Rogers, and R.E. Denton: Geophys. Res. Lett. **26**, 2163, 1999.
24. 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, P.L. Pritchett: J. Geophys. Res. **106**, 3715, 2001.
25. M. Hesse, J. Birn, and M. Kuznetzova: J. Geophys. Res. **106**, 3721, 2001.
26. J. Birn and M. Hesse: J. Geophys. Res. **106**, 3737, 2001.
27. A. Otto: J. Geophys. Res. **106**, 3751, 2001.
28. M.A. Shay, J.F. Drake, B.N. Rogers, and R.E. Denton: J. Geophys. Res. **106**, 3759, 2001.
29. Z.W. Ma and A. Bhattacharjee: J. Geophys. Res. **106**, 3773, 2001.
30. P. Pritchett: J. Geophys. Res. **106**, 3783, 2001.
31. L. Yin, D. Winske, S.P. Gary, and J. Birn: J. Geophys. Res. **106**, 10,761, 2001.
32. P.L. Pritchett: J. Geophys. Res. **106**, 25,961 (2001).

33. M. Hesse, M. Kuznetsova, and J. Birn: J. Geophys. Res. **106**, 29,831 (2001).
34. H. Karamabadi, D. Krauss-Varban, N. Omidi, and H.X. Vu: J. Geophys. Res. **104**, 12,313 (1999).
35. Z.W. Ma and L.C. Lee: J. Geophys. Res. **106**, 25,951 (2001).
36. J.D. Huba and L.I. Rudakov: submitted to Phys. Plasmas (2002).
37. J.A. Bittencourt, *Fundmentals of Plasma Physics* (Pergamon Press, New York, 1986).
38. J.D. Huba, Phys. Fluids B **3**, 3217 (1991).
39. A. Fruchtman and L.I. Rudakov: Phys. Rev. Lett. **69**, 2070 (1992).
40. A. Fruchtman and L. I. Rudakov: Phys. Rev. **E50**, 2997 (1994).
41. L.I. Rudakov and J.D. Huba: to appear in Phys. Rev. Lett. (2002).
42. J.D. Huba and J.G. Lyon: J. Plasma Phys. **61**, 391, 1999.
43. S. Zalesak: J. Comp. Physics **31**, 335 (1979).
44. K. Hain: J. Comput. Phys. **73**, 131, 1987.
45. C.K. Birdsall and A.B. Langdon: *Plasma Physics via Computer Simulation*, (Adam Hilger, New York 1991).

Fluid Plasma Simulation of Coupled Systems: Ionosphere and Magnetosphere

Antonius Otto[1] and Hua Zhu[1]

Geophysical Inst., Univ. Alaska Fairbanks, AK 99775, USA

Abstract. The coupling between the magnetosphere and the ionosphere is of major importance for the magnetospheric dynamics. The interaction of a highly collisionless plasma with a strongly collision dominated plasma requires an extension of the usual fluid plasma equations, to include the transport which dominates the ionospheric dynamics. Presented is a new simulation model which includes the ionospheric transport as source terms in the set of fluid plasma equations. Various aspects of the ionospheric influence are illustrated with the two-dimensional simulation of the formation and evolution of a field-aligned current layer in the ionosphere.

1 Introduction

The coupling between the magnetosphere and the ionosphere is of fundamental importance for all geomagnetic processes [10,9]. The magnetosphere is bound by the shocked solar wind and by the ionosphere on the Earthward side. There is a considerable Poynting flux of electromagnetic energy into the ionosphere particularly during times of geomagnetic activity. At the same time, the absorption of energy and momentum and the outflow of cold and heavy ions exert a considerable drag on magnetospheric dynamics by the ionosphere. It should be noted that in cases of particle precipitation a major portion of the plasma content of a magnetic flux tube is in the ionosphere.

In most global MHD models [see the corresponding articles in this volume] the ionosphere is treated as a partially conducting boundary with a height integrated conductance. Global thermospheric convection models [16,18] typically use a quasi-static approach where the ion inertia in the momentum equation is neglected. This, however, is valid only on time scales where the ion motion has relaxed into a local steady state and the approach eliminates the propagation of typical plasma waves such as Alfvén waves in the ionosphere. While these global approaches are perfectly reasonable on the large scales considered in the global models they are not valid for small spatial scales (km and sub-km scales) and fast temporal scales where the large scale models lack important physics. Figure 1 illustrates the many scales of auroral structure from hundreds of km's down to about 100 m width of discrete arcs with a separation of 5 to 10 km (bottom).

Small-scale filamentary structure is very frequent in the auroral ionosphere [2,5]. This is a strong indication for time-dependent processes which require ion inertia [13,12]. There is also strong evidence that small scale structure has a considerable influence on the global magnetospheric and ionospheric properties

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 193–211, 2003.

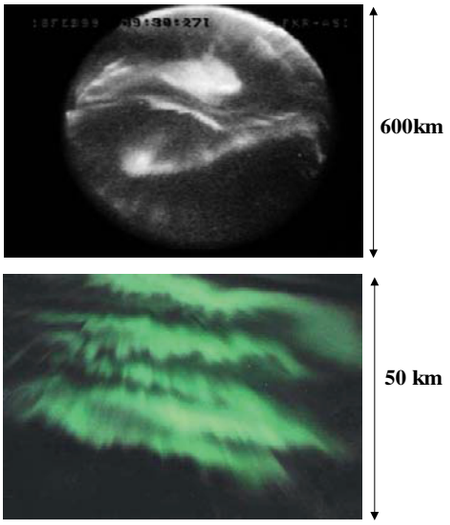


Fig. 1. Illustration of auroral arc scales in an all sky image (top, courtesy of Poker flat research range) and a higher resolution image(bottom)

such as the absorption of magnetospheric energy into neutral heating and the outflow of heavy ionospheric material which can exert a large drag on the magnetospheric dynamics. For instance, small scale ionospheric convection velocities (and the corresponding electric field) in the auroral region are often orders of magnitude larger than the large scale average fields. However, the energy absorbed by Joule heating is $\sigma_p \mathbf{E}_\perp{}^2$ with the Pedersen conductance σ_p and the convection electric field $\mathbf{E}_\perp$. Since large amplitude small scale fields are eliminated in the large scale averages, the large scale averaged fields will always underestimate the true Joule heating.

Following we will first introduce the basic equations, the numerical algorithm used to solve these equations, and discuss the basic physics associated with the ionospheric transport. Section 3 presents several simulation results to illustrate the capabilities of the model and Sect. 4 provides a discussion and summary.

2 Numerical Model

2.1 Basic Equations

The ionosphere connects the highly collisionless magnetospheric plasma with the collision dominated neutral atmosphere. Thus the ionospheric dynamics requires to include various effects which are not present or are unimportant in the highly collisionless plasma of the magnetosphere. The neutrals are a source of plasma in the case of ionization processes (e.g., due to energetic particle precipitation), they exert a frictional drag on the electrons and ions, and they provide a thermal

contact such the heated electrons and ions loose thermal energy to neutrals in proportion to the effective collision frequencies. Thus the equations need to include the coupling to the neutrals. In cases where the neutral dynamics plays a active role, a complete set of continuity, momentum and energy equations have to be included for the neutral species.

In summary the basic equations should consider the dominant transport terms which are ionization and recombination rates, ion-neutral, electron-neutral, and electron-ion collisions to determine the friction and energy exchange between the different species. Note that ionization and recombination not only appear in the continuity equations but also exchange momentum and energy between species. Last not least the collision frequencies depend on number densities and temperatures such that ion and electron energy (temperature or pressure) equations are needed.

There are two further effects which ought to be considered. The Hall effect is important in the ionosphere . In general Hall physics become important for gradients on the ion inertia scale c/ω_{pi} with the ion plasma frequency ω_{pi}. The ion inertia scale is between 1 and 10 km for typical ionospheric mass densities. Therefore Ohm's law (electron equation of motion) has to include the Hall term and the electron pressure term. The treatment of Hall physics in numerical modeling is further detailed in the corresponding chapter in this volume. In the ionosphere the Hall effect leads to the Hall currents in the lower E-region. Finally heat conduction from the magnetosphere is important for the electron energy budget and should be included in the electron energy equation. It is noteworthy that in addition to other terms the complete ion momentum equation should contain the ion inertia term. This term is usually neglected in global thermospheric convection models where the velocity is computed from the friction term in the fluid momentum equation.

The full set of equations consists of continuity, momentum, and energy equations for neutrals, ions, and electrons. Here we have assumed quasi-neutrality because the Debye length is very small. The magnetic field is computed from the the induction equation applied to generalized Ohms law. In the following equations the neutral equations present the sum over all individual neutral species and the same applies to the ion or bulk plasma equations. Thus the complete set of plasma-neutral equations [1] is given by

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \rho \boldsymbol{v} + \iota\rho - \beta\rho^2 \tag{1}$$

$$\frac{\partial \rho_n}{\partial t} = -\nabla \cdot \rho_n \boldsymbol{v}_n - \iota\rho + \beta\rho^2 \tag{2}$$

$$\begin{aligned}\frac{\partial \rho \boldsymbol{v}}{\partial t} &= -\nabla \cdot (\rho \boldsymbol{v}\boldsymbol{v}) - \frac{1}{2}\nabla p + \rho \boldsymbol{g} + (\nabla \times \boldsymbol{B}) \times \boldsymbol{B} \\ &\quad +\rho \boldsymbol{v}(\iota - \beta\rho) - \rho\nu_{in}(\boldsymbol{v} - \boldsymbol{v}_n)\end{aligned} \tag{3}$$

$$\begin{aligned}\frac{\partial \rho_n \boldsymbol{v}_n}{\partial t} &= -\nabla \cdot (\rho_n \boldsymbol{v}_n \boldsymbol{v}_n) - \frac{1}{2}\nabla p_n + \rho_n \boldsymbol{g} \\ &\quad -\rho \boldsymbol{v}_n(\iota - \beta\rho) - \rho\nu_{in}(\boldsymbol{v}_n - \boldsymbol{v})\end{aligned} \tag{4}$$

$$\frac{\partial \boldsymbol{B}}{\partial t} = \nabla \times \left(\frac{1}{ne}\nabla p_e\right) + \nabla \times (\boldsymbol{v} \times \boldsymbol{B}) - \nabla \times \left(\frac{1}{ne}(\nabla \times \boldsymbol{B}) \times \boldsymbol{B} + \eta \nabla \times \boldsymbol{B}\right) \tag{5}$$

$$\frac{\partial p}{\partial t} = -\boldsymbol{v} \cdot \nabla p - \gamma p \nabla \cdot \boldsymbol{v} + (\iota - \beta\rho) T_n + (\gamma - 1)\left[2\eta(\nabla \times \boldsymbol{B})^2 + \iota e_c - \frac{3\nu_{in}\rho}{m_i + m_n}(T_i - T_n) + \left(\frac{m_n \nu_{in}\rho}{m_i + m_n} + \iota m_i\right)(v - v_n)^2\right] + 3\nu_{\text{eff}}\rho(T_n - T_e) + \frac{\partial}{\partial z}\left(\lambda_e \frac{\partial T_e}{\partial z}\right) \tag{6}$$

$$\frac{\partial T_e^\kappa}{\partial t} = -\nabla \cdot (T_e^\kappa \boldsymbol{v}_e) + T_e^{\kappa-1}\left[3\nu_{ei}(T_i - T_e) + 3\nu_{\text{eff}}(T_n - T_e) + \frac{2\eta}{n}(\nabla \times \boldsymbol{B})^2 + \frac{\iota e_c}{n} - \frac{\iota e_c}{(\gamma - 1)n} + \frac{\partial}{n\partial z}\left(\lambda_e \frac{\partial T_e}{\partial z}\right)\right] \tag{7}$$

$$\frac{\partial p_n}{\partial t} = -\boldsymbol{v}_n \cdot \nabla p_n - \gamma_n p_n \nabla \cdot \boldsymbol{v}_n - (\iota - \beta\rho) k T_n + (\gamma_n - 1)\left[\frac{3\nu_{ni}\rho_n}{m_i + m_n}(T_i - T_n) - \iota\rho T_n + \frac{m_i \nu_{ni}\rho_n}{m_i + m_n}(v_n - v)^2\right] + 3\nu_{\text{eff}}\rho(T_e - T_n) \tag{8}$$

where ρ and ρ_n are the total plasma and neutral mass density; m_i and m_n are the ion and neutral particle mass; T_e, T_i, and T_n are the electron, ion, and neutral temperatures; p, p_e, and p_n are the total plasma pressure, electron pressure, and neutral pressure; n is the plasma number density; $\boldsymbol{v}$, $\boldsymbol{v}_e$, and $\boldsymbol{v}_n$ are plasma, electron, and neutral velocities; $\boldsymbol{B}$ is the magnetic induction vector; $\boldsymbol{g}$ is the gravitational acceleration; ι is the ionization rate; ν_{in} is the ion-neutral, ν_{ei} is the Coulomb, and ν_{eff} is the effective collision frequency between electrons and neutrals; η is the resistivity; γ and γ_n are the ratios of specific heats for the plasma and neutrals; e_c is the average energy that goes into electron heating for each ionization process (a typical value is 2 eV); λ_e is the electron heat conduction coefficient; and $\kappa = 1/(\gamma - 1)$. Energy conservation requires $\gamma_n = \gamma = 5/3$.

Ionization, recombination, and electron heating is parameterized by using input from an ionospheric transport computation [8]. The normalized resistivity considers electron-ion ν_{ei}, electron-neutral ν_{en}, and ion-neutral ν_{in} collision frequencies [17]

$$\eta = \eta_{ei} + \eta_{en} + \eta_{in} \tag{9}$$

with $\eta_{ei} = \lambda\nu_{ei}\tau_A$, $\eta_{en} = \lambda\nu_{en}\tau_A$, and $\eta_{in} = \lambda\nu_{in}\tau_A$ where ν_{ei}, ν_{en}, and ν_{in} are the electron-ion, electron-neutral, and ion-neutral collision frequencies; $\tau_A = l_0/v_A$ is the Alfvén time; and $\lambda = \left(\frac{c}{\omega_{pe} l_0}\right)^2$ is a normalization coefficient. The

electron heat conduction coefficient λ_e and the effective electron-neutral collision frequency ν_{eff} are from the TIME-GCM [16].

The equations assume: $n = n_e = n_i$, $\rho = n(n_e + m_i)$, $p = p_e + p_i$, and $\rho \boldsymbol{v} = \rho_e \boldsymbol{v}_e + \rho_i \boldsymbol{v}_i$.

Results of the simulation model will be displayed in physical units. However, it is useful to employ a suitable normalization to characteristic values (which is used in the basic equations (1) through (8)) where length scales are measured in units of a typical length $l_0 = 1.0$ km; plasma and neutral number density n and n_n are normalized to $n_0 = 50,000$ cm^{-3}; mass m_e, m_i, and m_n to the oxygen mass $m_0 = m_{O^+}$; mass density ρ and ρ_n to $\rho_0 = n_0 m_0$; magnetic field B to a typical horizontal magnetic field perturbation $B_0 = 100$ nT; velocity $\mathbf{v}$ to the Alfvén speed $v_A = B_0(\mu_0 \rho_0)^{-1/2}$ (2.44 km s^{-1}); pressure p, p_e, and p_n to $P_0 = B_0^2/(2\mu_0)$; T_e and T_n to $T_0 = P_0/(n_0 k)$ (k is the Boltzmann constant); and time t to characteristic Alfvén transit time $\tau_A = l_{z0}/v_a$ (0.4 seconds).

2.2 Numerical Algorithm

Most plasma fluid simulations consider a collisionless plasma where the coupling to a background component such as a neutral gas is small. Collisions between electrons and ions are negligible and resistive interaction is assumed to be limited to small spatial regions where turbulent wave particle interactions may be present. In these cases the basic equations can be cast into a conservative form yielding equations of the type

$$\frac{\partial f}{\partial t} = -\frac{\partial g}{\partial x} \tag{10}$$

where for simplicity we assumed a one dimensional form. Viscosity, resistivity, or heat conduction add a second derivative term. Further terms are added by ionization and recombination, frictional forces, and thermal contact involving different fluids and yield equations of the type

$$\frac{\partial f}{\partial t} = -\frac{\partial g}{\partial x} + \nu \frac{\partial^2 f}{\partial x^2} - \mu f \tag{11}$$

With $\nu = 0$ this partial differential equation (PDE) is hyperbolic. However, for $\nu \neq 0$ the equation becomes parabolic and for very large values of ν and μ the time derivative is negligible and the equation becomes an elliptic partial differential equation. The method for the numerical simulation has to be stable for different types of differential equations and sufficiently accurate to solve the system (1) to (8) accurately. Considering the complexity of these equations the method should be reasonably simple for the actual implementation. Particularly fully implicit methods as well spectral methods require the inversion of large and highly complex matrices [3]. Some common methods to solve hyperbolic fluid equations are discussed in the sections on global MHD simulations.

The method chosen for the simulation of the plasma-neutral equation uses the DuFort-Frankel scheme [14,11] with an additional iterative step to evaluate

the term μf. Explicitly we discretise this equation in the following form

$$\frac{f_i^+ - f_i^-}{2\Delta t} = -\frac{g_{i+1}^0 - g_{i-1}^0}{2\Delta x} + \nu \frac{f_{i+1}^0 - f_i^+ - f_i^- + f_{i-1}^0}{\Delta x^2} - (\mu f)_i^0 \tag{12}$$

where upper indices +, 0, and − indicate the time level of the integration (each separated by Δt), and lower indices correspond to the spatial grid indices. Here the time centered term $-2f_i^0$ is approximated through $-(f_i^+ + f_i^-)$ in the DuFort-Frankel scheme. Provided $\langle \mu f \rangle_i^0$ is available the full integration step becomes

$$f_i^+ = f_i^- + 2\Delta t H_i \tag{13}$$

with

$$H_i = \frac{1}{1+\alpha}\left[-\frac{g_{i+1}^0 - g_{i-1}^0}{2\Delta x} + \nu \frac{f_{i+1}^0 + f_{i-1}^0 - 2\nu f_i^-}{\Delta x^2} - (\mu f)_i^0 \right] \tag{14}$$

and $\alpha = 2\nu\Delta t/\Delta x^2$. A schematic of this integration scheme is shown in Fig. 2. However, μf is not available at time level 0 and grid index i. Therefore a two step process is applied with a first step in which $(\mu f)_i^0$ is evaluated from a spatial average and is corrected in the second step.

$$\widetilde{f}_i^0 = f_i^- + \Delta t H_i \left(\langle \mu f \rangle_i^0 \right) \tag{15}$$

$$f_i^+ = f_i^- + 2\Delta t H_i \left(\widetilde{\mu f}_i^0 \right) \tag{16}$$

The corresponding scheme is stable subject to the conditions

$$\Delta t < \min\left[\frac{u}{\Delta x}, \frac{1}{\mu} \right] \tag{17}$$

where $g = uf$ was assumed. For the actual scheme the velocity u should be replaced by a typical phase speed which is the fast mode speed for MHD applications and the maximum of fast mode and whistler speeds when the Hall term is included. Note that the maximum whistler phase and group velocities scale as

$$u = \frac{\omega}{k} \sim v_A \frac{\pi c}{\omega_{pi} \Delta x}, \tag{18}$$

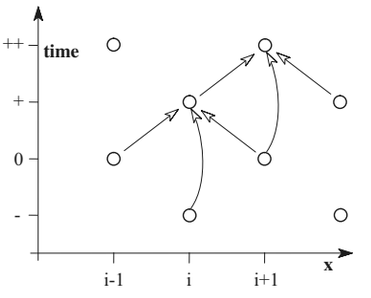


Fig. 2. Schematic of the integration scheme.

i.e., it is dependent on the grid resolution (see also the chapter on Hall Magnetohydrodynamics). For ionospheric applications typical collision frequencies are of the orders of a few 10^3 s^{-1} which usually dominate the term $u/\Delta x$. The stability condition is equivalent to the condition for physical accuracy which requires to resolve the propagation of typical waves and the fastest relaxation processes. The scheme is expected to provide second order accuracy. For ν, $\mu = 0$ the scheme reduces to the well known Leapfrog scheme. In addition we employ a small viscosity that is applied only locally if grid oscillations occur.

It should be noted that a detailed numerical stability analysis for the full set of basic equations is not available for several reasons. The usual stability analysis assumes a homogeneous background and considers linear perturbations. For most integration schemes single equations or simple sets of coupled equations are used in the stability analysis. A large system of coupled partial differential equations will generate a correspondingly large number of Eigenvalues which determine the stability. To our knowledge this task has not been carried out for the coupled plasma neutral equations. For nonlinear perturbations the numerical stability analysis has to be able to distinguish between physical and numerical instability. This problem is actually more complicated than the analysis of the physical instabilities, which is one of the reasons for performing a numerical simulation in the first place.

For the present case numerical properties are implied and tested using simplified equations or basic physics such as wave propagation, ionization, the deceleration of the plasma component due to the friction with the neutral gas, frictional heating, and the cooling through thermal contact. We will consider some of these aspects in the following sections.

2.3 Boundary Conditions

Periodic boundary conditions are used in the x direction, which allows waves to propagate freely through the system. At the top boundary of the system free boundary conditions are used for most quantities; that is, the value at the boundary is computed from its value inside the physical boundary of the system by extrapolation. This maintains the initial perturbation for the Alfvén waves and allows the transmission of the waves which are reflected in the lower ionosphere.

At the lower boundary of the simulation box (base of the ionosphere), the densities ρ, ρ_n, pressures p, p_e, p_n, and temperatures T_e and T_n, are given by continuous boundary conditions; that is, the boundary values of these quantities are computed by extrapolation. For the magnetic field $\mathbf{B}$, current $\mathbf{j}$, and plasma velocity $\mathbf{v}$, however, we determine the boundary values using Ohm's law.

The boundary values for ρ and p, are used to compute the collision frequencies ν_{en}, ν_{in}, and ν_{ei} and the Pedersen and Hall conductivities. The electric field is extrapolated from the physical boundary to a mathematical boundary and the Pedersen and Hall conductivities are employed to compute the current density from $\mathbf{j} = \sigma \cdot \mathbf{E}$ on this boundary. Similarly, the boundary value for the plasma

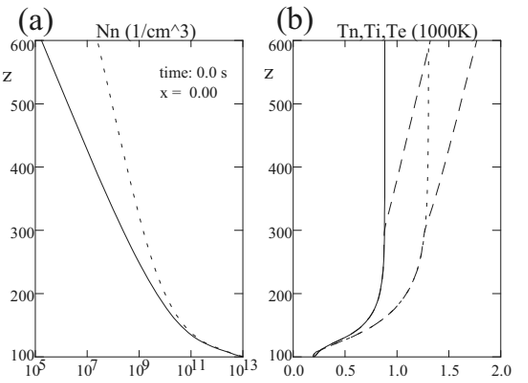


Fig. 3. (a) Altitude neutral density profile for solar minimum conditions (solid line) and solar maximum conditions (dotted line) conditions and (b) temperature profiles for the neutral atmosphere (solid line for solar minimum, dashed and dotted lines for solar maximum), for ions (initially identical to the neutrals), and for electrons (short dashes branching off the neutral temperature).

velocity $\mathbf{v}$ is determined by

$$\mathbf{v_i} = \frac{e}{m_i}\left(\frac{\nu_{in}}{\omega_{ci}^2+\nu_{in}^2}\mathbf{E}_\perp + \frac{\omega_{ci}}{\omega_{ci}^2+\nu_{in}^2}\mathbf{E}\times\hat{\mathbf{b}}\right), \tag{19}$$

where ω_{ci} is the ion gyro frequency and $\hat{\mathbf{b}}$ is the unit vector of unperturbed magnetic field. Finally, the magnetic field $\mathbf{B}$ is determined from the relation $\mathbf{j} = \nabla \times \mathbf{B}$.

2.4 Initial Configuration and Perturbation

The following applications of the plasma neutral simulation model employ an ionospheric region which extends from 100 km (lower E region) to 1100 km altitude with the main magnetic field in the vertical direction. We use Cartesian coordinates in the simulation with x and y (invariant) perpendicular to the main magnetic field (in the horizontal direction) and z is upward along the unperturbed magnetic field (which is directed downward in the northern hemisphere).

The grid is uniform in the horizontal (x) direction. In the vertical direction a non-uniform grid is used with a maximum resolution of 4 km at the base of the simulation box. The collision terms are only computed up to an altitude of 800 km. Above this the collisional effects are negligible for the considered time scales. Similarly the neutral dynamics is only required up to this altitude.

The neutral gas is initially in hydrostatic equilibrium with a temperature T_n and density n_n chosen for solar minimum and solar maximum conditions (Fig. 3) from the Mass Spectrometer Incoherent Scatter (MSIS) model [4,15].

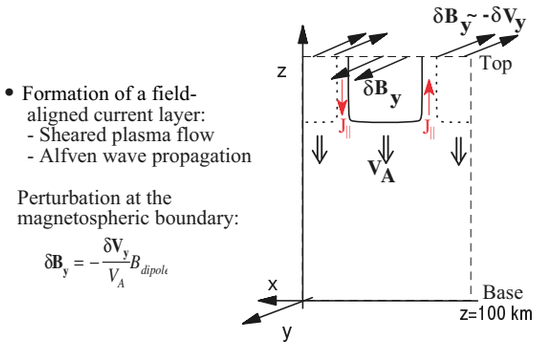


Fig. 4. Sketch of the initial perturbation

The neutral number density and temperature profiles are significantly different at high altitudes for solar minimum and solar maximum conditions.

An initial perturbation of the plasma velocity and the horizontal magnetic field is applied at the top boundary with

$$v_y = \frac{v_{y0}}{2} \left(1 + \tanh(0.015(z - 800))\right) \tanh(2.0(x + 6.0)) \tag{20}$$

for $x < 0$ and

$$v_y = -\frac{v_{y0}}{2} \left(1 + \tanh(0.015(z - 800))\right) \tanh(2.0(x - 6.0)) \tag{21}$$

for $x \geq 0$, and the magnetic field perturbation is given by

$$B_y = -v_y \sqrt{\mu_0 \rho}. \tag{22}$$

The perturbation defined in (9), (10), and (11) is part of the initial conditions and must be consistent with the boundary conditions. This initial perturbation propagates down into the ionosphere as a pair of Alfvén waves (Fig. 4) which produce two field-aligned current layers at $x = -6$ km (upward) and $x = 6$ km (downward). The normalized amplitude is $v_{y0} = 1.0$, which corresponds to 2.44 km s^{-1}.

The initial magnetic field perturbation is illustrated in the contour plot in Fig. 5 which shows the horizontal magnetic field perturbation (the perturbation is pointing into the plane) and the magnetic field-aligned electric current between the two Alfvén waves. The velocity perturbation has exactly the same shape as the magnetic field perturbation.

3 Results

3.1 Ion Heating

The initial magnetic and velocity perturbations propagate down into the ionosphere and are reflected in the lower F- and in the E-region (100 to 150 km

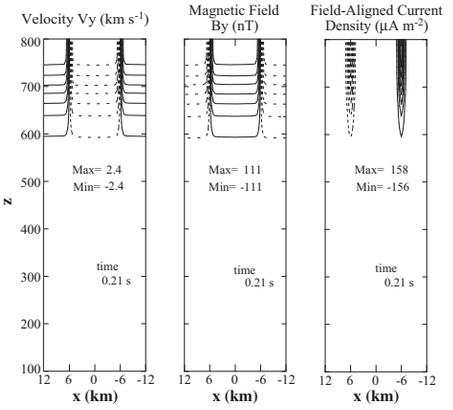


Fig. 5. Initial magnetic field, velocity, and field-aligned current density

altitude). The reflection occurs at about 0.6 seconds into the simulation and Fig. 6 shows the waves just after the reflection from the lower ionosphere. The physical cause for the reflection is twofold. The increasing ion number and mass density in the lower ionosphere generates a partial reflection because of the higher inertia of the plasma. However, most importantly the very large ion-neutral collision frequency in the lower ionosphere does not allow the ions to move freely. The Alfvén wave is not able to accelerate the ions to a velocity which maintains the amplitude and is thus reflected.

On large scales the reflection is necessary because an approximate steady state requires

$$J_p = \frac{\delta B_y}{\mu_0} = \Sigma_p B_z \delta v_y \tag{23}$$

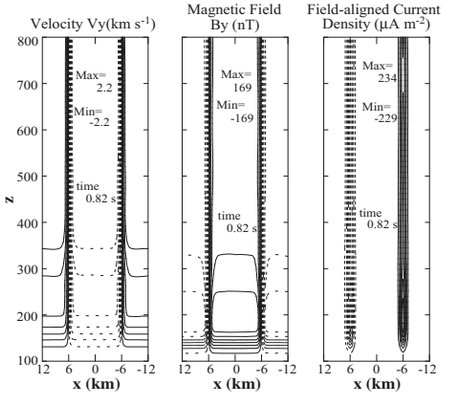


Fig. 6. Magnetic field, velocity, and field-aligned current density short after the reflection of the Alfvén waves from the lower ionosphere

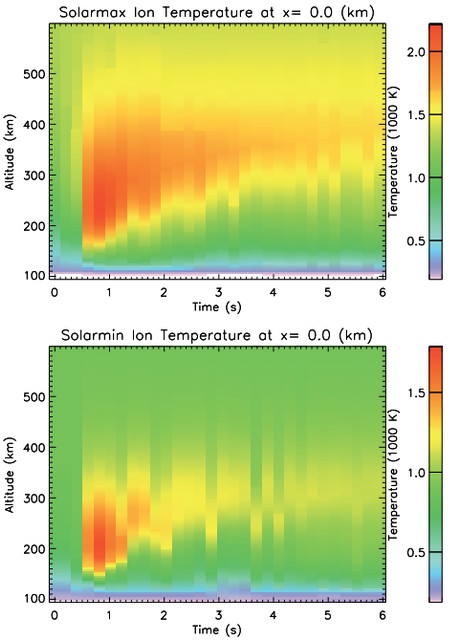


Fig. 7. Ion heating through the Alfvén wave for different solar conditions

where J_p and Σ_p are the height integrated Pedersen current and conductivity, δB_y and δv_y are the magnetic and velocity perturbation in the top ionosphere, and B_z is the dipole magnetic field component. Thus the reflected wave is needed to adjust the perturbation to be consistent with a steady state $\delta B_y/\delta v_y = \mu_0 \Sigma_p B_z$. The corresponding reflection coefficient for the Alfvén wave is

$$r = \frac{\mu_0 v_{Az} \Sigma_p - 1}{\mu_0 v_{Az} \Sigma_p + 1} \tag{24}$$

with v_{Az} being the Alfvén speed based on the dipole field.

The interaction with the lower ionosphere has a number of interesting aspects. The plasma motion associated with the Alfvén wave causes enhanced plasma neutral friction that heats the plasma. This heating is largest in the region where the collisions are highest and at the time when the plasma velocity is fastest.

Figure 7 illustrates the evolution of the ion temperature as a function of time and height at $x = 0$ for solar maximum and solar minimum conditions. Solar maximum conditions use a neutral temperature of about 1500° compared to about 800° for solar minimum conditions. Correspondingly the neutral atmosphere is more expanded and of higher density at greater altitudes thus leading to higher ion-neutral collision frequencies. The resulting friction heats the plasma therefore at greater heights and in a more extended region during solar maximum conditions. The plot illustrates that in both cases ion heating does not occur until the wave has reached the lower ionosphere. This plasma is cooled

through the thermal contact to the neutrals. In the absence of ionization and heat conduction (both of which are small for this case) the ion temperature is determined from equations (6) and (7) as

$$\frac{\partial T_i}{\partial t} = -\frac{\nu_{in} m_i}{m_i + m_n}(T_i - T_n) + \frac{1}{3k_B}\frac{\nu_{in} m_n}{m_i + m_n}(v - v_n)^2 \tag{25}$$

Thus both the heating rate through the velocity source term and the cooling rate through thermal contact with the neutral are proportional to the ion-neutral collision frequency. After the initial impact of the Alfvén wave the velocity settles to an approximate steady state and the ions assume time asymptotically a constant temperature

$$T_i - T_n = \frac{m_n}{3m_i}(v - v_n)^2 \tag{26}$$

The e folding time to relax to the steady temperature can be obtained by a cut a constant altitude in Fig. 7. It is fastest at lower altitudes because it is proportional to the ion-neutral collision frequency. Note that the ion temperature evolution also depends significantly on the altitude profile of the ion velocity which is established in this interaction. For the considered case the decay time for the temperature from equation 25 agrees well (within 5%) with our model results.

The velocity in the upper ionosphere does not decrease to zero after the reflection because we have assumed a long wave train (as would be realistic) and the velocity at greater heights is the superposition of the incoming and the reflected wave.

3.2 Current Closure and Density Depletion

An interesting aspect of the evolution is the closure of the field-aligned electric current. At the leading edge of the propagating Alfvén waves, the field-aligned current is closed by a perpendicular polarization current. After the reflection the magnetic perturbation is enhanced compared to the incoming Alfvén wave, such that two field-aligned current layers are formed with their base in the lower ionosphere as illustrated in Fig. 8. Note that the figure shows only the region from 100 km to 300 km altitude for better resolution.

The field-aligned current of these layers is closed by the Pedersen current. This current is carried by ions as illustrated in Fig. 8. In the lower ionosphere the ion-neutral collision frequency is larger than the ion gyro frequency such that the ions do not gyrate freely and are deflected perpendicular to the magnetic field in the direction of the electric field. However, the electron collision frequency is smaller than the gyro frequency such that electrons are still carrying out the $\mathbf{E} \times \mathbf{B}$ drift motion which points into the plane (parallel to the velocity perturbation of the incoming and reflected Alfvén waves). Therefore the ions carry the Pedersen current that closes the parallel currents and the electron Hall current is in the perpendicular direction.

The field-aligned current is carried by electrons because of the small inertia of the electrons (in this simulation electron inertia is in fact neglected) as shown

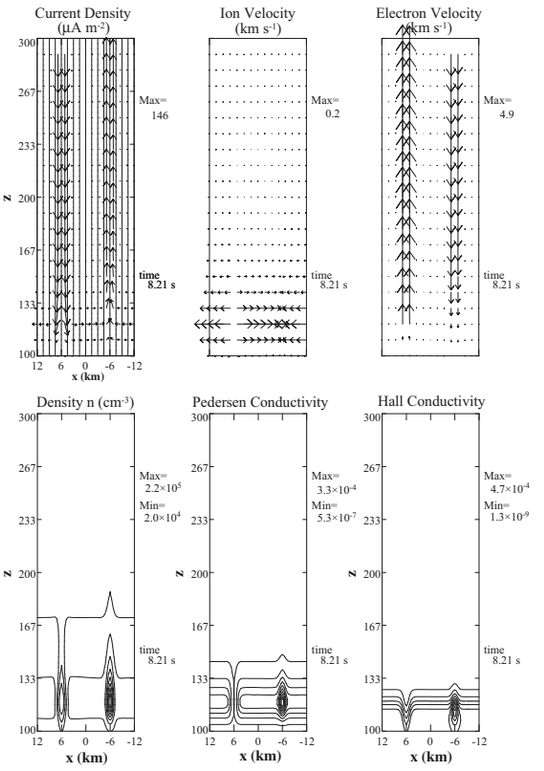


Fig. 8. Current density (top left), bulk or ion velocity (top middle), electron velocity (top right), plasma density (bottom left), Pedersen conductivity (bottom middle), and Hall conductivity (lower right)

in the top right plot of Fig. 8. The fact that the parallel current, which is carried by electrons, is closed by the ion Pedersen current has an interesting implication. Because of charge neutrality the divergence of the current density is zero $\nabla \cdot \boldsymbol{j} = 0$. This, however implies that $\partial n/\partial t = \nabla \cdot n\boldsymbol{v_e} = \nabla \cdot n\boldsymbol{v_i} \neq 0$. Thus the number density and the mass density has to change at the base of field-aligned current layers in the ionosphere. The density is increasing in layers with an upward current and it is decreasing in layers with a downward current. The mechanism is illustrated in Fig. 9.

The effect is quantitatively demonstrated in Fig. 10 for the converging and diverging currents at the base of the two current layers and outside the current layers. Within 8 seconds the density increases by almost a factor of 2 at the base of the upward current layer (top in Fig. 10) and it decreases by 85% at the base of the downward field-aligned current (bottom plot). The decreasing density outside the current layers is caused by recombination and is provided as a reference. Note the different color scale in each plot in Fig. 10.

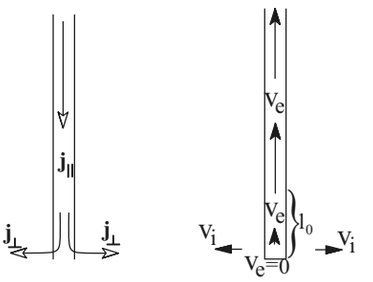


Fig. 9. Sketch illustrating the divergence and convergence of electron and ion currents

In this example we have imposed a field-aligned current of about $140\,\mu\mathrm{A/m^2}$ yielding an electron velocity of close to 5 km/s in the main field-aligned current layers. Using the continuity equation and a scale of about 20 km for the gradient in the electron velocity yields a time scale of a few seconds for changes in the density consistent with the results shown in Fig. 10. Thus field-aligned current can play a very significant role to modify the ionospheric density (in addition to precipitation or local instabilities). It is also worth pointing out that the modification in the electron density leads to a corresponding change in the

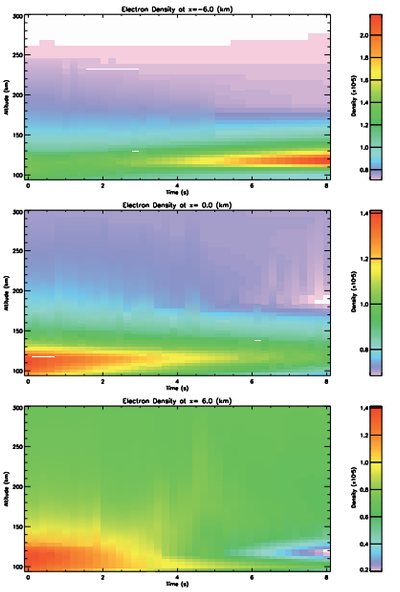


Fig. 10. Electron number density in the upward (top), downward (bottom) current layers and outside the field-aligned current

conductivities (Fig. 8). These in turn have an influence on the reflection of Alfvén waves and thus on the field-aligned current structure. Figure 8 demonstrates that the up- and downward currents are not anymore symmetric after only 8 seconds into the evolution. The change in the electron density lead to a widening of the upward current channel and a narrower downward current.

The simulation also allows to compare the currents which are directly computed from $\nabla \times \mathbf{B}$ with the usual Hall $\boldsymbol{j_h} = \sigma_h \widehat{\mathbf{b}} \times \mathbf{E}$ and Pedersen $\boldsymbol{j_p} = \sigma_p \mathbf{E}_\perp$ currents, e.g. [15], where $\widehat{\mathbf{b}}$ is the unit vector along the dipole field where we use the electric field from equation (4) and

$$\sigma_p = \frac{en}{B}\left(\frac{\nu_{en}\omega_{ce}}{\omega_{ce}^2+\nu_{en}^2} + \frac{\nu_{in}\omega_{ci}}{\omega_{ci}^2+\nu_{in}^2}\right)$$
$$\sigma_h = \frac{en}{B}\left(\frac{\omega_{ce}^2}{\omega_{ce}^2+\nu_{en}^2} - \frac{\omega_{ci}^2}{\omega_{ci}^2+\nu_{in}^2}\right)$$

The resulting currents agree excellent with the currents directly computed from the magnetic field perturbation in the region below about 200 km. Above this region the inertia terms (and possibly also pressure gradients) become important which are neglected in the derivation of Hall and Pedersen conductivities.

3.3 Ionization and Current Layer Dynamics

Thus far we have not addressed the effects of ionization caused by particle precipitation. The simulation model allows to switch on electron precipitation in pre-defined regions at the boundary. Depending on the neutral atmosphere, the precipitating energy flux, and the characteristic energy, ionization rates are computed by including parameterized results from energetic particle transport computations [8]. Note that while the precipitating flux carries some current, this current is often negligible and most of the field-aligned currents at ionospheric heights appear to be carried by thermal electrons.

Precipitation of energetic particle has a number of effects. First it will result in ionization which, for intense aurora, can change the local number density by an order of magnitude in seconds. The secondary collisions will cause significant heating of the electron fluid (which is cooled through the thermal contact to the neutrals) and the increasing number density will cause a different reflection of Alfvén waves and thus change the dynamics of field-aligned current generation. To illustrate this Fig. 11 shows the result of a very simple computer experiment with our model.

Here we use the same configuration as in the previous results but the model switches on precipitation with an energy flux of 40 mW/m^2 and a characteristic energy of 400 eV. However, this precipitation is switched on only in the region from $-10\,km < x < -2\,km$ with no precipitation (and ionization) outside of this region. Figure 11 show the resulting field-aligned current layers, the ion velocity, and the electron velocity in the entire region (top). The field-aligned current is again carried by the electron component and the Pedersen current by ions.

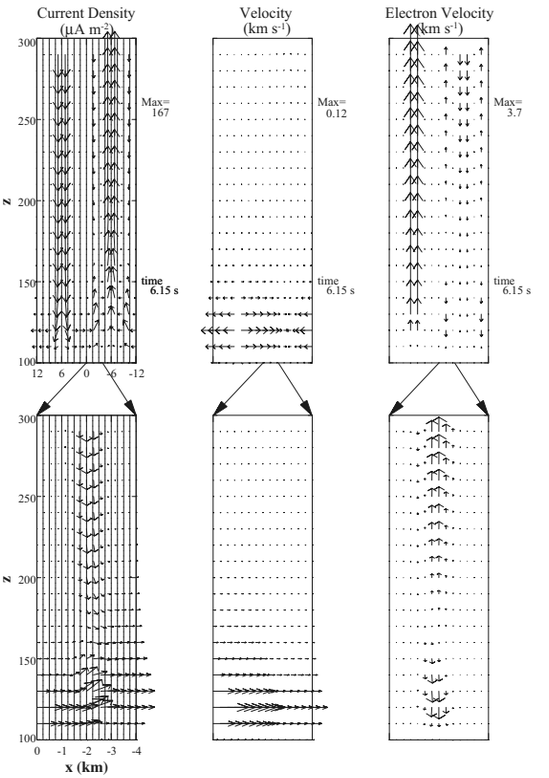


Fig. 11. Current density, ion velocity, and electron velocity in the entire simulation region (to a height of 300 km) and magnified for the region from $x = -4$ to $x = 0$ (bottom)

The top plot demonstrates two additional thin current layers in the vicinity of the upward current. The two additional current layers are located exactly at the precipitation boundaries. The current in the additional filaments is directed downward at higher altitudes and upward at the base of the ionosphere. This odd behavior has a rather simple explanation.

The ionization increases the number density in the precipitation region and thus also the Pedersen conductivity as illustrated in Fig. 12. This is for instance also expressed by the lower electron velocity in the main upward current layer which is due to the higher density thus maintaining the number flux and current at a fairly constant level. In the region of higher density the magnitude of the total Pedersen current increases and the current spreads over a larger extent in height.

Therefore the Pedersen current in the unperturbed region has to be deflected upward (through a field-aligned component carried by electrons) to close into the upward field-aligned current (which has a now a higher base). In addition

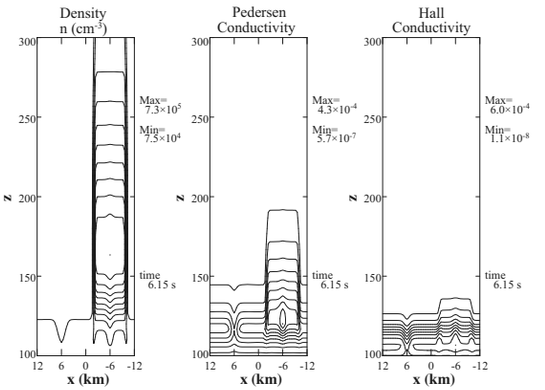


Fig. 12. Plasma density, Pedersen, and Hall conductivity

the larger total Pedersen current requires to draw some downward field-aligned current at the precipitation boundaries.

Note that this interpretation uses largely the concept of Pedersen currents which is strictly not valid on small spatial and temporal scales. However, the mechanism can be formulated in terms of Alfvén wave reflection with the same results. For instance the region where the up and downward component of the filamentary current diverges is located at the maximum in number density.

4 Discussion

Presented are a few relatively simple examples for small scale magnetosphere-ionosphere coupling simulations. It is important to emphasize that all of the examples require the propagation of Alfvén waves which in turn requires the ion inertia term. Without this term the the usual wave solutions for the fluid plasma are not present and it is not possible to study the formation of field-aligned current layers in the ionosphere.

In our first example we presented the ion and Joule heating as a result of Alfvén wave reflection. This example is a convenient test for the source terms of the set of plasma-neutral equations. Joule heating is an important aspect for the neutral dynamics and it can provide a pressure gradient to force an ion up-flow. In global models Joule heating is usually underestimated in regions of high auroral activity. The probable cause is the large scale averaging that is implied by global models which basically eliminates the large amplitudes (electric field, plasma velocity) of the small scale fluctuations.

The second example illustrates the formation of field-aligned current layers. It is demonstrated that strong field-aligned current layers can lead to a rapid modification of the ionospheric densities at the base of these current layers. These modifications like strong density gradients can cause local instabilities, change the local conductivities and are thus a source for ionospheric structure.

The density modifications also provide a nonlinear feedback to the system because they alter the way in which Alfvén waves are reflected and thus alter the field-aligned current structure. An aspect which we omitted is the strong electron heating in the field-aligned current. While energetically less important, this heating can provide a significant vertical pressure perturbation and may significantly contribute to upward ion acceleration.

The interaction of precipitation and Alfvén wave reflection can contribute to the generation of ionospheric structure and filamentary currents. While precipitation may not carry a major amount of field-aligned currents, thin precipitation boundaries are likely to be very important for field-aligned formation.

Finally it is worth to note that although we have used the picture of Hall and Pedersen currents the assumption to derive the corresponding conductivities may not be valid on small temporal and spatial scales in particular for F-region dynamics where the dynamic time scales can compete with the collision time scales such that plasma inertia becomes important.

Results for an actual auroral event [6] using this simulation code provide strong evidence for the presence of strong field-aligned current in the vicinity of discrete aurora [19].

Acknowledgments

This work is supported by the High Frequency Active Auroral Research Program (HAARP), under Air Force contract F94030e-d to the University of Alaska Fairbanks. The computation is supported by the Arctic Region Supercomputer Center at the University of Alaska, Fairbanks.

References

1. G. T. Birk, A. Otto, J. Comput. Phys. **125**, 513 (1996)
2. J. E. Borovsky, J. Geophys. Res. **98**, 6101, (1993)
3. Fletcher, Computational Techniques for Fluid Dynamics, I and II, Springer, 1988
4. A. E. Hedin, J. Geophys. Res., **96**, 1159, 1991.
5. B. S. Lanchester, M. H. Rees, D. Lummerzheim, A. Otto, H. U. Frey, and K. U. Kaila, J. Geophys. Res. **102**, 9741 (1997)
6. B. S. Lanchester, M. H. Rees, D. Lummerzheim, A. Otto, K. J. F. Sedgemore-Schulthess, H. Zhu, and I. W. McCrea, J. Geophys. Res. **106**, 1785 (2001)
7. D. Lummerzheim, and J. Lilensten, Ann. Geophys., **12**, 1039, 1994.
8. D. Lummerzheim, Rep. UAG-R-318, Geophys. Inst., Univ. of Alaska, Fairbanks, 1992.
9. G. Lu, A. D. Richmond, B. A. Emery, and R. G. Roble, J. Geophys. Res. **100**, 19,643 (1995)
10. R. L. Lysak, Space Sci. Rev. **52**, 33 (1990)
11. A. Otto, Comput. Phys. Commun. **59**, 185 (1990)
12. A. Otto, Advances in Space Research **18**, No. 8, 187 (1996)
13. A. Otto, and G. T. Birk, Phys. Fluids **B4** (11), 3811 (1992)
14. D. E. Potter, *Computational Physics*, (John Wiley, New York 1973)

15. M. H. Rees, Physics and Chemistry of the Upper Atmosphere, Cambridge Univ. Press, New York, 1989.
16. R. G. Roble, Solar Terrestrial Energy Program (STEP) Handbook of Ionospheric Models, edited by R. W. Schunk, Utah State University, Logan, UT, 281, 1996.
17. R. W. Schunk, Solar-Terrestrial Physics: Principles and Theoretical Foundations, Proceedings of the Theory Institute, Chestnut Hill, D. Reidel, Dordrecht, 609, 1983.
18. J. J. Sojka, Rev. Geophys. **7**, 371, 1989.
19. H. Zhu, A. Otto, D. Lummerzheim, M. H. Rees, and B. S. Lanchester, J. Geophys. Res. **106**, 1795 (2001)

Global Magnetohydrodynamics – A Tutorial

Joachim Raeder

Institute of Geophysics and Planetary Physics, University of California, Los Angeles, USA

Abstract. Global modeling and simulation of the complex Earth's space environment requires thorough understanding of the physical processes, the application of robust and sophisticated numerical techniques, an efficient implementation of the numerical algorithms, including parallelization, and comprehensive evaluation against data. This tutorial and review article introduces and discusses the very foundation of global modeling: the choice of numerical grids, the governing equations, numerical algorithms, error estimates, boundary conditions, magnetosphere-ionosphere coupling, and coupling with a thermosphere-ionosphere model. Two examples, simulations of a magnetospheric substorm and of a magnetic storm, show the utility as well as the limitations of the model and exemplify the current state-of-the-art and the lessons learned during the past decade. This article should help non-specialists to understand what goes into such models, what their main use is, and where global models have limitations. Those who are more familiar with global and large-scale models will find a review of latest results.

1 Introduction

Global modeling of geospace, that is, of the magnetosphere-ionosphere-thermosphere system, began about 20 years ago with the first simple magnetohydrodynamic (MHD) models of the solar wind - magnetosphere interaction [42,48]. It is thus a relatively young discipline compared to, for example, the modeling of the atmosphere. However, over this comparatively short period enormous progress has been made. While the first models were two-dimensional, it was soon realized that the magnetosphere is intrinsically three-dimensional, and such models appeared soon thereafter [18,55]. The next big step was the inclusion of electrodynamic ionosphere models that provided the closure of field-aligned currents (FACs) and the connection between magnetospheric and ionospheric convection [23,38,60,68,81]. These model extensions followed the realization that the ionosphere might, at least in part, control magnetospheric convection, and thus the magnetospheric dynamics in general. At this stage, the models were largely used to reproduce the large scale magnetospheric morphology and to investigate basic physical processes. However, it was not clear whether or not the model results and the underlying assumptions of magnetohydrodynamics were even correct, although they seemed to give correct answers for some parameters, for example the bow shock and magnetopause standoff distance. The ISTP program brought the first direct comparisons of model results with *in situ* measurements [25,26,65]. Since then there has been a flurry of model-data comparisons; too many to mention them all here. However, there have been several activities in which systematic comparisons were made, in particular the NSF/GEM convection challenge

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 212–246, 2003.

(see [50] and other articles in that issue) and the NSF/GEM substorm challenge (see [66] and other articles in that issue.) These studies have been particularly useful to assess the capabilities and limitations of current state-of-the-art models, that is, they (and other studies) have shown that global geospace models have now truly predictive capabilities. However, they have also shown deficiencies in many areas that require continued model development. Nevertheless, the models have become sufficiently efficient and sophisticated that they are now considered for use in the applications branch of space science – operational space weather forecasting.

In the following we discuss the physical and numerical foundation of global modeling, the coupling of different regions and processes, and examples of global simulations of substorm and storm events.

2 Global Modeling – Things to Consider

Ideally, the magnetosphere-ionosphere system should be modeled using the Vlasov equations, i.e., the collisionless Boltzmann equations for the plasma species together with Maxwell's equations for the fields. Because this is impractical given the limited computer resources, the magnetohydrodynamic (MHD) equations are commonly used as the basis of a numerical model. Although the MHD equations appear relatively simple at a first glance, the development of efficient and accurate numerical algorithms for their solution is a formidable task that has not yet come to conclusion. For magnetosphere modeling the problem is exacerbated by the presence of multiple spatial and temporal scales. Therefore numerous different approaches have been developed, and the existing models differ in many respects. However, a number of issues is common to all models, and we discuss in the following their relevance and the advantages and disadvantages of various approaches.

2.1 Simulation Geometry and Numerical Grids

Any large-scale simulation starts with the choice of a numerical grid and the associated numerical methods. In the case of global simulations of Earth's magnetosphere the simulation boundaries should be well within supermagnetosonic flows, i.e., generally $\geq 18\ R_E$ from Earth on the sunward side, $\geq 200\ R_E$ in the tailward direction, and $\geq 50\ R_E$ in the transverse directions. However, these values are only a guide. For example, if the solar wind magnetosonic Mach number is very low the bow shock can move to several 10's of R_E upstream, in which case the sunward boundary must be farther away from Earth to keep the bow shock within the simulation domain.

There is a variety of choices for numerical grids. However, none of them is optimal and they all have their distinct advantages and disadvantages. Consequently, virtually all possible grid approaches have been used to some extent.

Uniform Cartesian grids, like the one shown in Fig. 1a provide lowest programming overhead, lowest computing overhead, essentially no memory overhead, easiest parallelization, and near perfect load balancing for parallelized

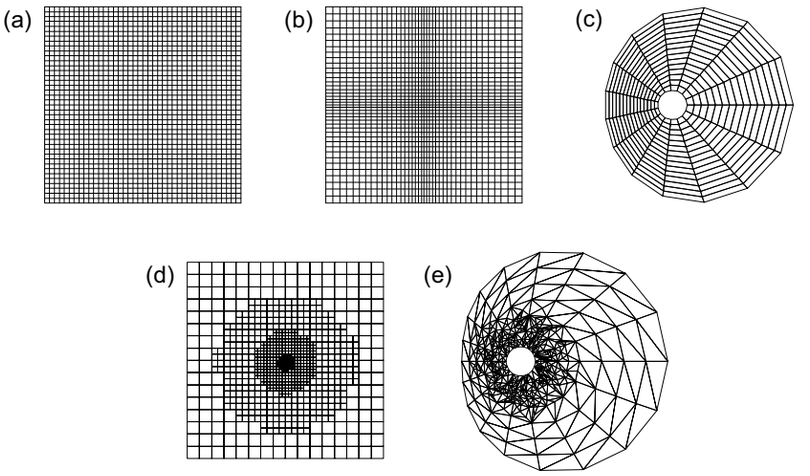


Fig. 1. Several common choices for numerical grids: **(a)** a uniform Cartesian grid, **(b)** a stretched Cartesian grid, **(c)** a non-Cartesian grid with Cartesian topology, **(d)** a structured adaptive grid, **(e)** a unstructured grid

computation [55,88,56]. The major drawback is that such grids are not adapted to the solution. Consequently computational resources are wasted where they are not needed (in regions where the solutions are smooth) while other regions are under resolved, for example, sharp gradients and shocks.

Stretched Cartesian grids like the one shown in Fig. 1b can be better adapted to the solution, while maintaining essentially all of the advantages of a uniform Cartesian grid [60,83]. In the case of global magnetospheric simulations such a grid can actually be quite well adapted, providing high resolution in the X (sun-Earth) direction at the bow shock and the magnetopause, high resolution in the Z direction in the tail plasma sheet, and substantially lower resolution almost everywhere else. Consider a typical simulation box that is $300\times100\times100$ $R_E^3 = 3\times10^6$ R_E^3 large. At a uniform 0.25 R_E resolution such a grid would require 1.92×10^8 cells, whereas a stretched Cartesian grid can achieve a 0.25 R_E resolution in the critical parts of the magnetosphere with $\sim$1-2$\times10^6$ cells. Thus, a stretched Cartesian grid requires about two orders of magnitude less computational resources. Such a grid is used, for example, in the UCLA code [60,68,64].

Grids as shown in Fig. 1c (non-Cartesian, but with Cartesian topology) are irregular but still with a regular connectivity between grid cells. This allows the grid better to be adapted to the solution with only small overhead in computation, however, post-processing and visualization becomes significantly more difficult. The Lyon-Fedder-Mobarry (LFM) code [49,24] uses such a grid.

A relatively new gridding strategy is based on overlaying grid patches with increasingly smaller resolution, such as shown in Fig. 1d. This approach is often call "structured adaptive mesh refinement" (SAMR) when combined with the

dynamical adaptation of the grid to the solution, that is, grid patches are created and destroyed as the solution evolves in time [10,9]. The ratio of the gridsize between different *levels* in the grid hierarchy is a fixed integer (usually 2 or 4), and so is the timestep. Different refinement strategies are possible, like a block structure in which all patches are of the same size and ordered in a tree (for example, the BATS'R'US code of the Michigan group [59,29]). SAMR promises the most accurate solutions for a given number of grid cells. However, SAMR incurs substantial programming and computer overhead. Parallelization and load balancing of SAMR codes is extremely difficult.

Unstructured grids, like the one shown in Fig. 1e are suitable for finite element (FE) and finite volume (FV) methods. They are often constructed from triangles (in 2d) or tetrahedrons (in 3d), but other basic building blocks are also possible. Despite their geometrical flexibility, unstructured grids are rarely used in plasma simulation because of their high programming and computational overhead and because of the difficulty to parallelize such codes.

2.2 The Governing Equations

Although the magnetohydrodynamic (MHD) equations are often under scrutiny when applied to space plasmas, experience has proven that they are adequate in many situations where the spatial scale of interest is larger than the ion gyroradius and the ion inertial scales, and the temporal scale is longer than the ion gyroperiod. In assessing the validity of the MHD equations one must consider that they are *conservation equations*. Specifically, MHD describes the conservation of mass, momentum, energy, and magnetic flux. As far as the plasma is concerned the only significant underlying assumption is that the velocity distribution functions of the plasma constituents are only a function of $|\boldsymbol{v} - \boldsymbol{v}_d|$ in phase space, where $\boldsymbol{v}_d$ is the drift speed (first moment of the distribution). This is trivially fulfilled for a Maxwellian distribution and causes all moments higher than the scalar pressure to vanish (in the case of a Maxwellian distribution), or at least to decouple [8]. Violations of the $f(\boldsymbol{v} - \boldsymbol{v}_d)=f(|\boldsymbol{v} - \boldsymbol{v}_d|)$ assumption are mostly mild in large parts of the magnetosphere. However, they can be significant in the ring current, in regions of strong diffusion, and possibly in the plasma sheet. In places where such deviations from a symmetric distribution occur higher order moments, for example the heat flux tensor, come into play. The correct closure of the equation set becomes then becomes an issue [34,35]. In some cases it is possible to augment the MHD equations appropriately (for example by adding anomalous diffusion terms); however, in other cases (for example, the ring current) a different formalism is indicated [89].

The MHD equations can be written in different forms, which are all mathematically equivalent, but generally lead to different numerical methods. In the following, the symbols have their usual meaning, e.g., $\boldsymbol{B}$ and $\boldsymbol{E}$ are the magnetic and electric field, respectively, $\boldsymbol{v}$ is the plasma velocity, ρ is the density, p is the pressure, $\boldsymbol{j}$ is the current density, η is a resistivity, $\underline{\underline{\mathrm{I}}}$ is the unit tensor, and γ is the ratio of specific heats.

Non-conservative (Primitive Variable) Formalism

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \boldsymbol{v}) \tag{1a}$$

$$\frac{\partial \boldsymbol{v}}{\partial t} = -(\boldsymbol{v} \cdot \nabla)\boldsymbol{v} - \frac{1}{\rho}(\nabla p - \boldsymbol{j} \times \boldsymbol{B}) \tag{1b}$$

$$\frac{\partial p}{\partial t} = -(\boldsymbol{v} \cdot \nabla)p - \gamma p \nabla \cdot \boldsymbol{v} \tag{1c}$$

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

$$\nabla \cdot \boldsymbol{B} = 0 \tag{1e}$$

$$\boldsymbol{E} = -\boldsymbol{v} \times \boldsymbol{B} + \eta \boldsymbol{j} \tag{1f}$$

$$\boldsymbol{j} = \nabla \times \boldsymbol{B} \tag{1g}$$

The primitive variable formulation leads to numerical schemes that do not strictly conserve momentum and energy, even in the hydrodynamic case. Such schemes do not guarantee correct shock speeds and correct jump conditions at discontinuities [41]. Furthermore, the convective derivative $(\boldsymbol{v} \cdot \nabla)$ is difficult to treat numerically. Although the use of the primitive variable formulation leads to algorithms with low memory requirements, its use should be avoided because much better approaches are available.

Full Conservative Formalism

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \boldsymbol{v}) \tag{2a}$$

$$\frac{\partial \rho \boldsymbol{v}}{\partial t} = -\nabla \cdot \left\{ \rho \boldsymbol{v}\boldsymbol{v} + \underline{\underline{I}}(p + \frac{B^2}{2}) - \boldsymbol{B}\boldsymbol{B} \right\} \tag{2b}$$

$$\frac{\partial U}{\partial t} = -\nabla \cdot \left\{ (U + p)\boldsymbol{v} + \boldsymbol{E} \times \boldsymbol{B} \right\} \tag{2c}$$

$$\frac{\partial \boldsymbol{B}}{\partial t} = -\nabla \times \boldsymbol{E} \tag{2d}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{2e}$$

$$\boldsymbol{E} = -\boldsymbol{v} \times \boldsymbol{B} + \eta \boldsymbol{j} \tag{2f}$$

$$\boldsymbol{j} = \nabla \times \boldsymbol{B} \tag{2g}$$

$$U = \frac{p}{\gamma - 1} + \frac{\rho v^2}{2} + \frac{B^2}{2} \tag{2h}$$

The full conservative formulation allows the application of conservative finite difference schemes that strictly conserve mass (ϱ), momentum ($\varrho \boldsymbol{v}$), energy (U), and magnetic flux. This formulation is therefore always preferable. It may lead, however, to difficulties in low β regions ($\beta = p/(B^2\mu_0)$ is the ratio of the plasma pressure to the magnetic field pressure) where the pressure becomes the difference of two large numbers. Numerical errors can then cause nonphysical negative pressures. A semi-conservative form of the equations may then be more appropriate.

Gas Dynamic Conservative (Semi-conservative) Formalism

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \boldsymbol{v}) \tag{3a}$$

$$\frac{\partial \rho \boldsymbol{v}}{\partial t} = -\nabla \cdot (\rho \boldsymbol{v}\boldsymbol{v} + p\underline{\underline{\mathrm{I}}}) + \boldsymbol{j} \times \boldsymbol{B} \tag{3b}$$

$$\frac{\partial e}{\partial t} = -\nabla \cdot (\{e + p\}\boldsymbol{v}) + \boldsymbol{j} \cdot \boldsymbol{E} \tag{3c}$$

$$\frac{\partial \boldsymbol{B}}{\partial t} = -\nabla \times \boldsymbol{E} \tag{3d}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{3e}$$

$$\boldsymbol{E} = -\boldsymbol{v} \times \boldsymbol{B} + \eta \boldsymbol{j} \tag{3f}$$

$$\boldsymbol{j} = \nabla \times \boldsymbol{B} \tag{3g}$$

$$e = \frac{\rho v^2}{2} + \frac{p}{\gamma - 1} \tag{3h}$$

The semi-conservative formulation allows for difference schemes that numerically conserve of mass (ϱ), momentum ($\varrho \boldsymbol{v}$), and plasma energy (e), but with no strict conservation of total energy. On the other hand, low β regions pose no difficulty. This approach can be combined with a full conservative scheme by integrating both energy equations (2c) and (3c) and using a 'β switch', as suggested by Balsara and Spicer [7].

The above equations are normalized with arbitrary normalization factors (3 of them are independent.) This is possible because the MHD equations have no intrinsic length or time scale. This changes, however, when Earth's dipole is introduced, which essentially provides the normalization of the magnetic field.

Frequently the ideal (that is, non-dissipative) MHD equations need to be augmented with a term for anomalous resistivity η. While any numerical code produces numerical resistivity, enough to enable magnetic reconnection, in some circumstances this may not be sufficient. A notable example is the dynamical evolution of substorms [67]. Although the precise mechanisms that cause anomalous diffusion are not well known, it is generally believed that anomalous diffusion is a function of the local current density, thus a suitable parametrization is given by, for example:

$$\eta = \alpha {j'}^2 \quad \text{if} \quad j' \geq \delta, \quad 0 \text{ otherwise} \tag{4}$$

$$j' = \frac{|j| \Delta}{|B| + \epsilon} \tag{5}$$

where j' is a normalized local current density, Δ is the grid spacing, and δ and α are empirical constants [64].

2.3 Boundary and Initial Conditions

The outer boundary conditions of the simulation domain are relatively straightforward. However, the inner boundary, where the magnetospheric field meets the ionosphere, is much more involved.

Sunward Side. The boundary conditions can be either arbitrary (fixed or time dependent), or measured solar wind data can be used. However, solar wind data are usually limited to one (or at best a few) solar wind monitors, thus there is very little knowledge of the true 3d structure of the solar wind. This leads to a

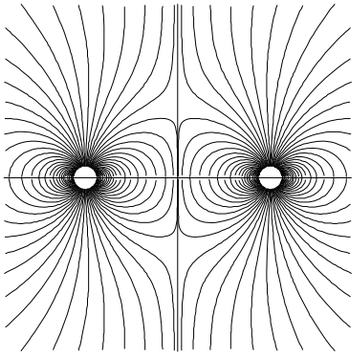


Fig. 2. Initial magnetic field constructed with a mirror dipole

problem with B_x: The three-dimensional structure of the solar wind needs to be known because

$$\nabla \cdot \boldsymbol{B} = 0 \qquad \Longleftrightarrow \qquad \boldsymbol{n} \cdot (B_{upstream} - B_{downstream}) = 0, \tag{6}$$

across any discontinuity in the IMF. This implies that $B_x = B_n$ cannot change if solar wind parameters are independent of Y and Z (simple extrapolation). A possible solution is to find the predominant normal vector $\boldsymbol{n}$ in the sense that all solar wind discontinuities during some time period are only a function of $\boldsymbol{n}$. This is difficult with a single solar wind monitor; however, boundary normal methods (for example, the minimum variance method [76,77]) can be applied.

On All Other Sides. Here, free flow conditions can be applied, i.e.,

$$\frac{\partial \Psi}{\partial n} = 0 \tag{7}$$

for all variables Ψ, except for the normal magnetic field which must be derived from the $\nabla \cdot \boldsymbol{B} = 0$ condition and which must be consistent with the numerical scheme.

Magnetic Field Initial Conditions. The magnetic field can be initialized by the superposition of dipole with mirror dipole to create $B_x = 0$ surface sunward of Earth (see Fig. 2). The field on sunward side is then replaced with the initial solar wind field, providing a $\nabla \cdot \boldsymbol{B} = 0$ transition.

Plasma Initial Conditions. The initial plasma conditions are usually given by a cold (5000 °K), tenuous (0.1 cm^{-3}), uniform plasma. From the start of the simulation it takes about 0.5-1 hour real time for the magnetosphere to form. However, the magnetosphere can have a substantial memory of prior conditions (possibly many hours), thus it is advisable to provide at least a few hours lead time from the start of a simulation up to a specific event.

2.4 MHD Numerics

Time Differencing. Consider the model equation:

$$\frac{\partial U}{\partial t} = -\nabla \cdot \boldsymbol{F}(U) \tag{8}$$

which is representative for the plasma part of the conservative or semi-conservative MHD equations (2,3).

Fairly simple difference schemes can be applied to the time derivative with second order accuracy (that is, numerical errors are proportional to Δt^2; more on that later), for example, the explicit predictor-corrector scheme:

$$U^{n+\frac{1}{2}} = U^n - \frac{1}{2}\Delta t \nabla \cdot \boldsymbol{F}(U^n) \qquad , \qquad U^{n+1} = U^n - \Delta t \nabla \cdot \boldsymbol{F}(U^{n+\frac{1}{2}}) \; , \tag{9}$$

or the explicit leap-frog scheme:

$$U^{n+1} = U^{n-1} - 2\Delta t \nabla \cdot \boldsymbol{F}(U^n, U^{n-1}) \; . \tag{10}$$

These schemes are generally accurate enough, sufficiently simple, and require moderate storage. However, they suffer from a stability requirement (the Courant-Friedrichs-Levy, or CFL criterion [75]) that limits the stable time step to Δt_{max}:

$$\Delta t_{max} \leq \delta \frac{\min(\Delta x, \Delta y, \Delta z)}{|\boldsymbol{v}| + v_{MS}} \; , \tag{11}$$

where δ is a constant of the order $O(1)$. The CFL criterion can be very restrictive because $\Delta t < \Delta t_{max}$ must be satisfied everywhere in the simulation domain, not just locally. In some parts of the magnetosphere the alfvén speed can become very large, severely limiting the stable time step. It is possible to apply the 'Boris correction' [13,17] or some variant thereof which limits the alfvén speed. This essentially entails the simultaneous reduction of the $\boldsymbol{J} \times \boldsymbol{B}$ and the perpendicular (to the magnetic field) component of the ∇p force in regions where the alfvén speed would be too high. A judicious choice of the reduction factor allows for much larger time steps without any adverse effects on the solutions.

Implicit time differencing schemes, where the right hand side involves variables at time level $n+1$:

$$U^{n+1} = U^n - \Delta t \nabla \cdot \boldsymbol{F}(U^{n+1}, U^n, U^{n-1}, \ldots) \tag{12}$$

can be unconditionally stable, but generally require the solution of large linear systems, which is computationally very expensive and generally impractical.

Spatial Discretization. The spatial discretization of the MHD equations is much more difficult than the time discretization. There are basically four different approaches: a) finite differences (FD), b) finite volume (FV) methods, which usually reduce to FD methods on Cartesian grids, c) finite element (FE) methods, and d) spectral methods.

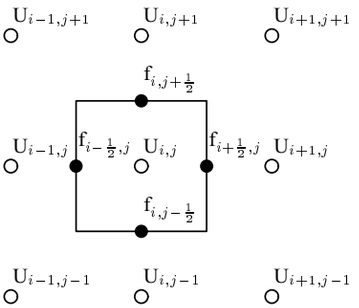


Fig. 3. Variable placement of the numerical fluxes

Since FD methods are most widely used in global magnetosphere models we restrict the discussion to these in the following. However, many FD concepts carry over to the other methods as well. More specifically, we focus on the discussion of *conservative difference schemes* since these are most suitable for global simulations.

Consider again the model equation, but this time with emphasis on the right hand side:

$$\frac{\partial U}{\partial t} = -\nabla \cdot \boldsymbol{F}(U) \tag{13}$$

where U is some variable and $\boldsymbol{F}(U)$ the flux associated with that variable. Introduce a regular Cartesian grid (in 2d) where the cell centers are at $x_i = i\Delta x, i = 1, \ldots$ and $y_j = j\Delta y, j = 1, \ldots$ and the cell corners are at $(x_{i+1/2,j+1/2}, y_{i+1/2,j+1/2})$. Discretize the right hand side of equation (13) as:

$$\frac{\partial U}{\partial t} = -(f_{i+\frac{1}{2},j}(U) - f_{i-\frac{1}{2},j}(U))/\Delta x - (f_{i,j+\frac{1}{2}}(U) - f_{i,j-\frac{1}{2}}(U))/\Delta y \;, \tag{14}$$

where we introduced the *numerical fluxes* $f_{i+\frac{1}{2},j}$ and $f_{i,j+\frac{1}{2}}$, which are functions of the grid values:

$$f_{i+\frac{1}{2},j} = G_x(\ldots, U_{i-1,j}, U_{i,j}, U_{i+1,j}, \ldots) \tag{15}$$

$$f_{i,j+\frac{1}{2}} = G_y(\ldots, U_{i,j-1}, U_{i,j}, U_{i,j+1}, \ldots) \;, \tag{16}$$

and which must be *consistent* with the physical flux $\boldsymbol{F}(U)$ in the following sense:

$$\boldsymbol{G}(U, \ldots, U, U, \ldots, U) = \boldsymbol{F}(U) \;. \tag{17}$$

Writing equation (13) in integral form:

$$\frac{\partial}{\partial t} \int_V U dV = \int_S \boldsymbol{F} d\boldsymbol{s} \;, \tag{18}$$

where V is an arbitrary, simply connected volume, S its surface, and $\boldsymbol{s}$ its surface normal, it is now easy to see that the variable U is globally conserved. In Fig. 3

this volume is taken to be one cell (i, j). Considering the cell that contains $U_{i,j}$, any change of $U_{i,j}$ must be opposite to the sum of the changes of the four surrounding cell variables $U_{i-1,j}$, $U_{i+1,j}$, $U_{i,j-1}$, $U_{i,j+1}$ caused by the flow of U through the cell boundaries. Thus, the sum of $\partial U_{i,j}/\partial t$ over the entire grid is zero, except for the fluxes through the physical boundaries.

Using the discretization (14) only guarantees the global conservation of the quantity U. The accuracy of the approximation is determined by the construction of the numerical fluxes $f_{i+\frac{1}{2},j}$ and $f_{i,j+\frac{1}{2}}$.

A few popular schemes are listed in the following; for simplicity we drop the second dimension:

- The second order central scheme:

$$f_{i+\frac{1}{2}} = \frac{1}{2}(F(U_i) + F(U_{i+1})) \tag{19}$$

- The fourth order central scheme:

$$f_{i+\frac{1}{2}} = \frac{7}{12}(F(U_i) + F(U_{i+1})) - \frac{1}{12}(F(U_{i-1}) + F(U_{i+2})) \tag{20}$$

- The Lax scheme:

$$f_{i+\frac{1}{2}} = \frac{1}{2}(F(U_i) + F(U_{i+1})) - \frac{1}{2}(U_{i+1} - U_i) \tag{21}$$

- The two step Lax Wendroff scheme: Use Lax scheme for predictor, and second order central for corrector.
- The Rusanov scheme:

$$f_{i+\frac{1}{2}} = \frac{1}{2}(F(U_i) + F(U_{i+1})) - \frac{1}{4}(|v_i| + |v_{i+1}| + c_i + c_{i+1})(U_{i+1} - U_i) \tag{22}$$

  where c is the sound speed.
- The Godunov type schemes solve a Riemann problem (i.e. the decay of a step function into waves) at the cell interface $i + \frac{1}{2}$ and compute the fluxes directly from the wave propagation (see e.g. [92,20] and references therein.) This is very accurate for gas-dynamics, but difficult for MHD because the system of equations involves a degenerate eigenvector related to $\nabla \cdot \mathbf{B} = 0$ [19].

Error Terms. The error terms associated with the spatial discretization can be found via a Taylor series expansion:

$$\begin{aligned} \Delta x \frac{\partial U}{\partial t} = &-(f_{i+\frac{1}{2}} - f_{i-\frac{1}{2}}) + a_1(\Delta x)^2 \frac{\partial^2}{\partial x^2} F(U) + b_1(\Delta x)^3 \frac{\partial^3}{\partial x^3} F(U) \\ &+ a_2(\Delta x)^4 \frac{\partial^4}{\partial x^4} F(U) + b_2(\Delta x)^5 \frac{\partial^5}{\partial x^5} F(U) + \dots \quad . \end{aligned} \tag{23}$$

The coefficients a_1, b_1, etc. depend of course on the choice of the numerical flux scheme. The error terms associated with even derivatives cause numerical *diffusion*, that is, they tend to smear out the solution, in particular at discontinuities

where the derivatives of the solution become large. Conversely, the error terms associated with odd derivatives cause numerical *dispersion*, which manifests itself mostly by ripples and under/overshoots near discontinuities. Dispersion is in a certain sense the worst enemy, because it may lead to nonphysical solutions, such as negative density or pressure.

Ideally, one wants to construct numerical fluxes that minimize the error terms. The order of a scheme is defined as the smallest order of the derivative with non-vanishing coefficient minus one, for example, a third order scheme will only have error terms proportional to the fourth and higher derivatives of the solution. First order schemes are thus primarily diffusive (with second order diffusion terms). All symmetric central schemes have no diffusion at all, that is, all a terms vanish. Although that may seem a desirable property, the dispersion of these schemes makes them virtually useless for any application that involves shocks, such as the magnetosphere, because of extreme over- and under-shoots near the shock and other discontinuities.

A more desirable property of the numerical scheme is therefore *monotonicity*. A scheme is called *monotone* if it lets no *new* extrema develop in the solution. Of course, there may be physically relevant extrema in the solution; however, one can show that, at least in 1d, the number of extrema is non-increasing.

Unfortunately, *Harten's Lemma* [32] states that a monotone scheme is at most first order accurate. Thus, the price for global monotonicity seems to be the large diffusion of first order schemes, such as the Lax or the Rusanov scheme. Harten (and others) also proposed the solution to this dilemma, which essentially amounts to the *hybridization* of the numerical fluxes. Instead of using one and the same scheme in the entire domain, a first order scheme is employed where the gradients in the solutions are large, while in those regions where the solution is smooth a higher order scheme is used. In its simplest form the numerical flux is then computed as follows:

$$f_{i+\frac{1}{2}} = \theta_{i+\frac{1}{2}} f^h_{i+\frac{1}{2}} + (1 - \theta_{i+\frac{1}{2}}) f^l_{i+\frac{1}{2}} \tag{24}$$

where f^l is a low order (for example Rusanov) flux, and f^h is a high order (for example second or fourth order central) flux [33,94]. The conservation properties of the scheme are preserved and do not depend on which fluxes are used and how they are combined. Note that flux hybridization is not possible for the non-conservative equations.

The switch function θ acts as a *Flux Limiter* and is generally a function of gradients in the solution, e.g., a function of the grid values surrounding $i + \frac{1}{2}$. The hybridization procedure is not always written in the above stated form and the term *Flux limiter* is often used in a somewhat different, but related context [37]. However, the principle is always the same. There is no optimal choice for a flux limiter, and numerous schemes have been developed. Most notable among them are the original hybrid method [33], *Flux Corrected Transport (FCT)* [14,93,94,21], *Total Variance Diminishing (TVD)* schemes [31,80,90,91], *Essentially Non-oscillatory (ENO)* schemes [43,39], and the *Van Leer* flux limited schemes [84,85,86].

Magnetic Flux Conservation. A particular difficulty of MHD simulations (as opposed to hydrodynamic simulations) is the conservation of magnetic flux, expressed as the Maxwell equation $\nabla \cdot \mathbf{B}$=0. $\nabla \cdot \mathbf{B}$=0 is an initial condition since $\nabla \cdot \mathbf{B}$ is conserved by Faraday's law:

$$\nabla \cdot \frac{\partial \boldsymbol{B}}{\partial t} = \frac{\partial (\nabla \cdot \mathbf{B})}{\partial t} = -\nabla \cdot \nabla \times \boldsymbol{E} = 0 \tag{25}$$

Most numerical schemes do not *a priori* preserve $\nabla \cdot \mathbf{B}$. For such schemes the accumulation of $\nabla \cdot \mathbf{B}$ can lead to serious errors, in particular spurious parallel acceleration, wrong magnetic topology (field lines that are not closed), and significant errors in the shock jumps [16,82]. There are a few methods to "clean" the magnetic field of monopoles, for example the projection method:

$$\nabla^2 \Psi = -(\nabla \cdot \mathbf{B}) \tag{26}$$

produces a monopole potential that can be used as a correction:

$$\boldsymbol{B}' = \boldsymbol{B} + \nabla \Psi \tag{27}$$

Note that the projection method requires the solution of a Poisson equation on the global grid which can be quite costly. Because the numerical solution of this equation has errors as well, the projection method can only achieve $\nabla \cdot \mathbf{B} = 0$ to a certain order in the gridspacing [82].

An alternative approach is to modify the MHD equations in such a way that $\nabla \cdot \mathbf{B}$ becomes a convected quantity [59,29]:

$$\frac{d(\nabla \cdot \mathbf{B})}{dt} = 0 \tag{28}$$

This approach does not guarantee any limit on the accumulation of $\nabla \cdot \mathbf{B}$ and may also lead to the violation of the shock jumps [82].

A preferable approach is to use a scheme that conserves magnetic flux *a priori*. Such a scheme was first introduced by Evans and Hawley [22] in the context of MHD simulations. Flux conservation is achieved by using staggered grids for the magnetic and electric field, such that the magnetic field components are placed on the center of cell faces:

$$(B_x)_{i+\frac{1}{2},j,k},\ (B_y)_{i,j+\frac{1}{2},k},\ (B_z)_{i,j,k+\frac{1}{2}}\ ,$$

and the electric field (the numerical flux for the $\boldsymbol{B}$ integration) on the centers of the cell edges:

$$(E_x)_{i,j+\frac{1}{2},k+\frac{1}{2}},\ (E_y)_{i+\frac{1}{2},j,k+\frac{1}{2}},\ (E_z)_{i+\frac{1}{2},j+\frac{1}{2},k}\ ,$$

as indicated in Fig. 4. The magnetic field time integration becomes then particularly simple, for example:

$$\frac{\partial}{\partial t}(B_x)_{i+\frac{1}{2},j,k} = \tag{29}$$
$$\{(E_y)_{i+\frac{1}{2},j,k+\frac{1}{2}} - (E_y)_{i+\frac{1}{2},j,k-\frac{1}{2}}\}/\Delta z$$

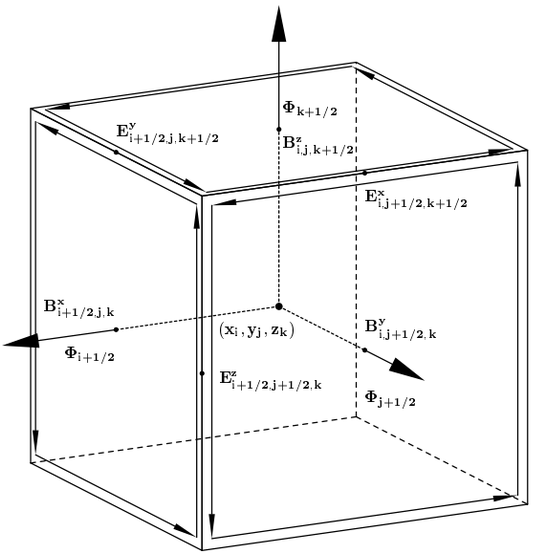


Fig. 4. Variable placement on a staggered grid for magnetic flux conserving integration of Faraday's law. The arrows along the cell edges indicate the electric field contributions in the magnetic field time integration step

$$-\{(E_z)_{i+\frac{1}{2},j+\frac{1}{2},k} - (E_z)_{i-\frac{1}{2},j+\frac{1}{2},k}\}/\Delta y \ .$$

and analogously for B_y and B_z. By advancing the field components in this way on all 6 cell faces and summing up it follows:

$$\frac{\partial}{\partial t}\int\int_{cell} \Phi df = \Delta y \Delta z (\frac{\partial B_x}{\partial t})_{i-\frac{1}{2}} + \Delta y \Delta z (\frac{\partial B_x}{\partial t})_{i+\frac{1}{2}} + \Delta x \Delta z (\frac{\partial B_y}{\partial t})_{j+\frac{1}{2}} + \ldots$$

$$= \{((E_y)_{i+\frac{1}{2},j,k+\frac{1}{2}} - (E_y)_{i+\frac{1}{2},j,k+\frac{1}{2}}) +$$

$$((E_y)_{i+\frac{1}{2},j,k-\frac{1}{2}} - (E_y)_{i+\frac{1}{2},j,k-\frac{1}{2}}) + \ldots\} \Delta x \Delta y \Delta z = 0 \ , \tag{30}$$

and thus the combined magnetic flux Φ through all 6 cell faces remains unchanged (Φ = const.) during the time integration, as required by equation (25). Note that the field can be divergence-free initialized by using a vector potential $\boldsymbol{A}$ in place of $\boldsymbol{E}$ [22].

Coordinate Transformation for Stretched Cartesian Grids. It is particularly simple to integrate the equations on a stretched Cartesian grid. Let the grid coordinates be given by analytic functions of the grid indices (i, j, k), that is: $x = x(i)$, y=$y(j)$, z=$z(k)$, then:

$$\frac{\partial}{\partial x} F(x,y,z) == \frac{\partial F}{\partial i}\frac{\partial i}{\partial x} = \frac{\partial F}{\partial i}\left(\frac{\partial x}{\partial i}\right)^{-1} \ , \tag{31}$$

and analogously for y and z derivative. The derivatives on the regular equidistant (i,j,k) grid need then only be multiplied with the appropriate geometric factors:

$$\frac{\partial}{\partial x}F(x,y,z) = \frac{\partial F}{\partial i}\left(\frac{\partial x}{\partial i}\right)^{-1} \tag{32}$$

$$\frac{\partial}{\partial y}F(x,y,z) = \frac{\partial F}{\partial j}\left(\frac{\partial y}{\partial j}\right)^{-1} \tag{33}$$

$$\frac{\partial}{\partial z}F(x,y,z) = \frac{\partial F}{\partial k}\left(\frac{\partial z}{\partial k}\right)^{-1}. \tag{34}$$

3 Coupling of Different Regions and Processes

Modeling the magnetosphere extends beyond solving the MHD equations. At a minimum, the MHD model needs an inner boundary at which FACs generated in the magnetosphere close through the resistive ionosphere. This process is commonly implemented such that the MHD calculation only extends to within 3-4R_E from Earth. Within that boundary the FACs are mapped along dipole field lines into the ionosphere. At the ionosphere end a potential equation is solved on a sphere (or a section thereof) to yield the ionospheric convection potential [23]. The potential is then mapped back to the inner boundary of the MHD calculation where it is used as boundary condition for the flow and field integration ($\boldsymbol{v} = (-\nabla\Phi) \times \boldsymbol{B}/|B|^2$). This mapping is illustrated in Fig. 5.

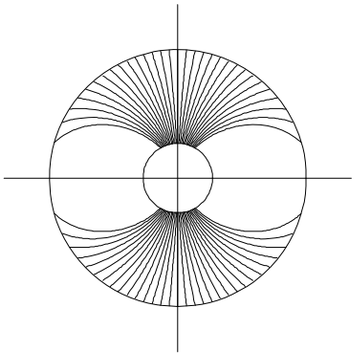


Fig. 5. Schematic of the field line mapping between the magnetosphere and the ionosphere

Such a mapping typically covers latitudes from $\sim 58^0$ to 90^0. In the region between the ionosphere and the inner boundary of the magnetosphere the MHD equations are not solved. This is partly necessitated by the high alfvén speeds in this region, and also by the fact that the relevant processes on these field lines

are not well described by the MHD equations, but are for the most part of a kinetic nature.

The relevance of the ionosphere for magnetospheric dynamics is best explained by the limiting cases. In the case of a vanishing ionospheric potential Φ, equivalent to infinite ionospheric conductance, the electric field and the convection velocity will also vanish ($\Phi = 0 \longrightarrow \boldsymbol{E} = 0 \longrightarrow \boldsymbol{v} = 0$). Thus, there is no convection in the ionosphere and field lines are tied. Ultimately (on a time scale of less than one hour) magnetospheric convection has to cease as well. In the opposite case the conductance of the ionosphere is zero which means that no current can flow from the magnetosphere through the ionosphere ($j_\parallel \to 0$). In that case field lines slip free through the ionosphere and the Earth (because the solid Earth's conductivity is small). Thus, magnetospheric convection can proceed uninhibited.

In reality the ionosphere has a finite conductance and field lines are dragged through the ionospheric plasma, dissipating energy that must be supplied from the magnetosphere via Poynting flux [58,78]. Thus, the ionosphere influences magnetospheric convection, and the primary controlling factor is the ionospheric conductance.

The magnetosphere-ionosphere coupling can be described either in a mechanical way, that is by calculating the stresses and motions of the ionospheric constituents (ions, electrons, and neutrals [79]), or by treating it as an electric circuit [87]. The latter approach is far easier to implement. Field-aligned currents are calculated at the magnetospheric boundary and used as input to the ionospheric potential equation. The polar ionosphere can be treated as a 2d shell to a very good approximation (because field lines are nearly radial), thus:

$$\nabla \cdot \underline{\underline{\Sigma}} \cdot \nabla \Phi = -j_\parallel \sin I \tag{35}$$

with the boundary condition Φ=0 at the magnetic equator. Because the ionosphere is a magnetized and partially ionized plasma the ionospheric conductance is a tensor [79], given by:

$$\underline{\underline{\Sigma}} = \begin{pmatrix} \Sigma_{\theta\theta} & \Sigma_{\theta\lambda} \\ -\Sigma_{\theta\lambda} & \Sigma_{\lambda\lambda} \end{pmatrix} \tag{36}$$

$$\Sigma_{\theta\theta} = \frac{\Sigma_P}{\sin^2 I} \quad , \quad \Sigma_{\theta\lambda} = \frac{\Sigma_H}{\sin I} \quad , \quad \Sigma_{\lambda\lambda} = \Sigma_P \tag{37}$$

where Σ_H is the Hall conductance, Σ_P is the Pedersen conductance, θ is the magnetic latitude, λ is the magnetic longitude, and I is the magnetic field inclination.

The potential calculation requires the specification of the ionospheric Hall and Pedersen conductance. These can either be assumed to be uniform (not a good assumption), be computed using empirical formulations, or be computed using a full-fledged ionosphere-thermosphere model.

The conductances are proportional to the ionospheric electron density (essentially E-region), which in turn is primarily given by solar EUV irradiance and

precipitation of magnetospheric electrons. The conductance contribution of the former can be easily parameterized from measurements, for example as [54]:

$$\Sigma_H = F_{10.7}^{0.53}(0.81\cos\chi + 0.54\cos^{1/2}\chi) \tag{38}$$

$$\Sigma_P = F_{10.7}^{0.49}(0.34\cos\chi + 0.93\cos^{1/2}\chi) \tag{39}$$

where $F_{10.7}$ is the solar radio flux (used as a proxy for solar EUV radiation) and χ is the solar zenith angle. Magnetospheric electron precipitation is either diffuse (pitch angle scattering of hot magnetospheric electrons), or discrete (accelerated auroral electrons.) The former can be parameterized by:

$$F_E = n_e(kT_e/2\pi m_e)^{\frac{1}{2}} \qquad , \qquad E_0 = kT_e \tag{40}$$

where T_e and n_e are the magnetospheric electron temperature and density, respectively. F_E is the energy flux, and E_0 is the mean energy of the precipitating electrons. Discrete electron precipitation can be modeled using the Knight relation [40]:

$$\Delta\Phi = K\max(0, -j_\parallel) \tag{41}$$

$$K = \frac{e^2 n_e}{\sqrt{2\pi m_e kT_e}} \tag{42}$$

$$F_E = \Delta\Phi_\parallel j_\parallel \qquad , \qquad E_0 = e\Delta\Phi_\parallel \tag{43}$$

where $\Delta\Phi$ is the parallel potential drop on an auroral field line.

The Pedersen and Hall conductances can then be computed using empirical formulas, for example [73]:

$$\Sigma_P = [40E_0/(16 + E_0^2)]F_E^{1/2} \tag{44}$$

$$\Sigma_H = 0.45E_0^{5/8}\Sigma_P \tag{45}$$

or by feeding the precipitation parameters, along with the potential, into a large-scale ionosphere-thermosphere model which then computes the conductances self-consistently from the electron-neutral collisions. The latter approach has recently been taken by combining the UCLA magnetosphere-ionosphere (MI) code with the NOAA Coupled Thermosphere Ionosphere Model (CTIM, [28,69]).

The CTIM part of the coupled model is a global multi-fluid model of the thermosphere–ionosphere system with a long heritage [28]. CTIM solves both neutral and ion fluid equations self-consistently from 80 to 500 km for the neutral atmosphere and from 80 to 10,000 km for the ionosphere on a spherical grid with 2° latitude resolution and 18° longitude resolution. The thermosphere part solves the continuity equation, horizontal momentum equation, energy equation, and composition equations for the major species O, O_2, and N_2 on 15 pressure levels. The ionosphere model part solves the continuity equations, ion temperature equation, vertical diffusion equations, and horizontal transport for H^+ and O^+, while chemical equilibrium is assumed for N_2^+, O_2^+, NO^+, and N^+. The

horizontal ion motion is governed by the magnetospheric electric field. The coupled model includes about 30 different chemical and photo-chemical reactions between the species. CTIM's primary inputs are the solar UV and EUV fluxes (parameterized by the solar 10.7 cm radio flux), the tidal modes (forcing from below), auroral precipitation, and the magnetospheric electric field, each of which is usually taken from parameterized empirical models. CTIM provides several outputs that are of prime importance for space weather, for example, global two- and three-dimensional ionosphere and thermosphere state fields, like electron density, neutral density, neutral wind, chemical composition, NmF2, hmF2, and total electron content (TEC). A more thorough description of CTIM, including the detailed equations, reaction rates, and examples can be found in [28].

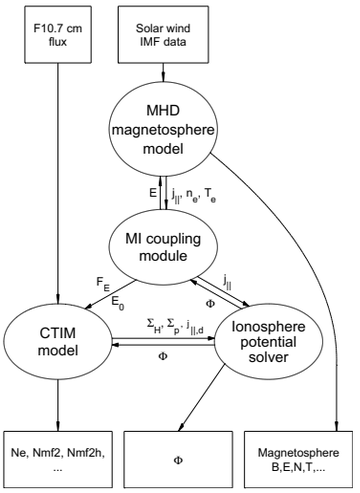


Fig. 6. Schematic showing the coupling between the UCLA magnetosphere-ionosphere model with the NOAA Coupled Thermosphere Ionosphere Model (CTIM)

The coupling of the models is schematically shown in Fig. 6. The MI model provides the electron precipitation parameters and the magnetospheric field-aligned currents (FACs.) In turn, CTIM provides the ionospheric conductance and the ionospheric dynamo current to the MI model. Thus, as far as the MI model is concerned, we replace empirical conductance calculations [73] that was used in most prior studies with first-principle calculations and also account for the ionospheric dynamo effect. The latter effect is probably of minor importance in most situations, but may become significant during storm recovery [72]. With this coupling, CTIM is also driven with more realistic magnetospheric inputs and depends on fewer empirical parameters.

4 Examples: Substorms and Storms

An important aspect of any kind of numerical modeling is to ensure the correctness of the model. Here, two issues come into play. First, one needs to verify that the model solves the underlying equations correctly. However, because numerical models compute only *approximations* to the underlying equations, the issue rather becomes to quantify the approximation errors. Ideally, this can be done by comparing model results with known analytic solutions. For global models, however, this approach has limitations because there are no known analytic solutions that even come close to the true complexity of the magnetosphere. Thus, much simpler analytic tests are usually conducted, for example, shock tube problems, magnetic field convection problems, equilibrium solutions, and checks of the conservation properties of the code (mass, momentum, energy, and magnetic flux). Such tests give some idea of how well an algorithm works, however, because of their simplicity they still leave open the possibility of serious errors in the simulation of the complex magnetosphere. The second issue involves the validity of the underlying equations themselves. Here, limited comparisons can be made with more sophisticated local models. For example, one can compare reconnection geometries and reconnection rates with those predicted by kinetic models. However, such comparisons are problematic because it is virtually impossible to set up test cases in a way that the boundary conditions for either model are the same. On the other hand, global models can be run with measured solar wind and IMF data as input and their results can then be compared with *in situ* observations. By doing so, both of the two issues above are addressed at the same time. If the simulation results match the *in situ* data well one can be confident that the equations are sufficient to describe the phenomena *and* that the numerical solution is sufficiently accurate. Inevitably, there will also be differences found between the simulation results and the data. Careful analysis of these differences, together with our theoretical understanding of the physical processes involved, will usually show if they are the result of numerical errors or the result of deficiencies in the underlying equations and assumptions. For example, the lack of ring current formation in global magnetosphere models is clearly caused by the simplifications inherent in the MHD equations which neglect the drift physics of the more energetic plasma populations. On the other hand, excessive reconnection rates may be the result of numerical effects like limited resolution or numerical diffusion.

Global geospace modeling is therefore most powerful if done in conjunction with data analysis. Very often it is impossible to draw solid conclusion from single (or a few) spacecraft observations. For example, a spacecraft may observe some key signatures of reconnection in the tail; however, that is not proof that reconnection actually occurs because other processes could produce indistinguishable signatures. This is the essence of current debates about the nature of substorms [45,53,6,5]. On the other hand, a simulation alone, showing reconnection, proves nothing because the simulation may simply be wrong for a variety of reasons, for example bad input data, lack of resolution, too much dissipation, or missing physics. A typical example is the question of whether or not the magnetotail

closes during extended periods of northward IMF [61,30,62]. Here, different models give different answers, and only a study that includes both simulations and data can be convincing. Furthermore, it is essential to test the limits of any given model, and the only possible ground truth are the observations. Global models distinguish themselves here from local models because they are actually testable by running them with measured input data and comparing the results with *in situ* observations. Local models, on the other hand, are virtually not testable because the boundary conditions are not precisely known.

That said, we compare two simulations – run with measured solar wind input – and compare the results with data. These two simulations address key issues in magnetospheric research: substorms and storms.

4.1 Substorms

The substorm of November 24, 1996, with onset at ∼2230 UT was chosen as a "GEM challenge" event because it was a typical isolated substorm following an extended period of magnetospheric quiet [66].

Figure 7a shows the IMF and solar wind data for the period of interest. After an extended period of northward IMF the IMF turns southward at around 2100 UT. Over the following 90 minutes the substorm growth phase commences until the expansion phase onset occurs around 2230 UT. It is not clear from the data (see below) whether the onset is triggered by the IMF northward turn [51] around this time; the simulation shown here indicates that this is not the case.

Figure 7b shows ground magnetometer data from the IMAGE chain compared with results from the simulation. The simulation captures the essence of the onset; however, the growth phase electrojets are stronger in the simulation, the onset is weaker, and the expansion phase is shorter.

Figure 8A compares, among other things, the polar cap magnetic flux from the simulation with estimates from Polar VIS data. Both the simulation and the data show an increase during the growth phase; however, the data indicate a larger saturation value. The beginning of the decrease coincides roughly with the expansion phase onset.

Figure 8B compares the IMP-8 magnetic field observations from the middle tail (around (-36,-3,10) R_E in the northern lobe) with the simulation results. The lobe magnetic field is a good indicator for flux and energy storage in the tail. Although the results agree qualitatively, there are significant quantitative differences. First, the loading is delayed in the simulation, and second, it is weaker than in the observed values. This indicates that there are (despite other correlations that look good) still significant discrepancies in the tail dynamics during the growth phase, similar to the ones observed in the growth phase electrojets. On the other hand, the unloading looks strikingly similar, except for a small time delay in the simulation.

Figure 9a shows the comparison with Geotail data. Geotail was also located in the middle tail ((-25,-8,-3) R_E in GSE coordinates) but close to, or in the plasma sheet. Since, whenever a spacecraft is close to a discontinuity or a sharp gradient, even a small displacement of the spacecraft can cause large changes

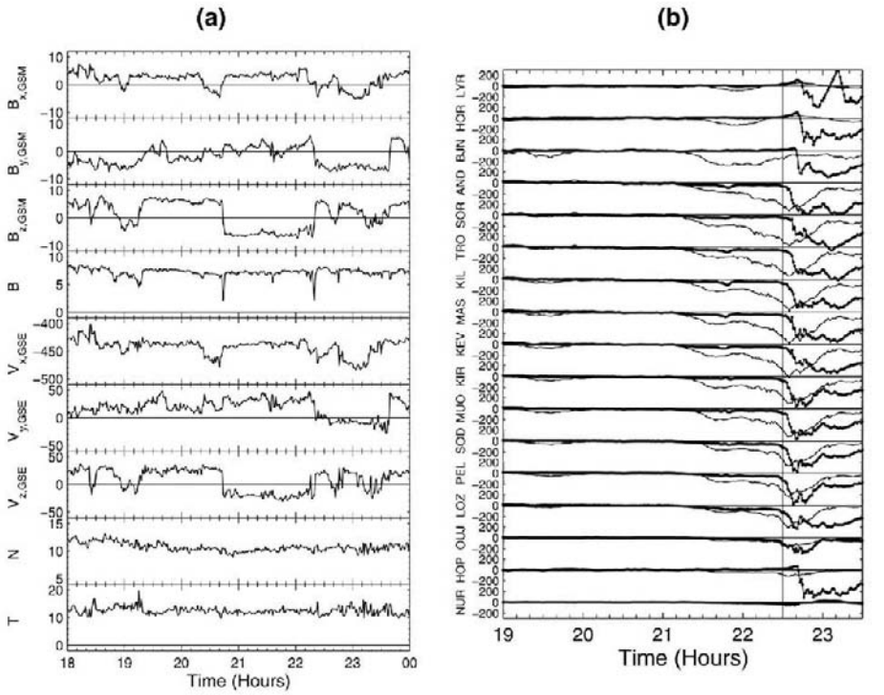


Fig. 7. **(a)** Wind interplanetary magnetic field (IMF) and solar wind data from (73,-18,8) R_E GSE on November 24, 1996. From top to bottom: the magnetic field components B_x, B_y, and B_z; the total magnetic field (all in nT, GSE); the flow velocity components V_x, V_y, and V_z (in km s^{-1}, GSE); the number density (in cm^{-3}); and the temperature (in eV). **(b)** Ground magnetometer traces from the International Monitor for Auroral Geomagnetic Effects (IMAGE) magnetometer chain on November 24, 1996 (thick dotted lines), and the comparison with the model result (thin solid line). Shown is the north-south (X) component in units of nT. The stations are Longyearbyen (LYR), Hornsund (HOR), Bear Island (BJN), Andenes (AND), Sørøya (SOR), Tromsø (TRO), Kilpisjärvi (KIL), Masi (MAS), Kevo (KEV), Kiruna (KIR), Muonio (MUO), Sodankylä (SOD), Pello (PEL), Lovozero (LOZ), Oulujärvi (OUJ), Hopen Island (HOP), Nurmijärvi (NUR)

in the observed values, we bracket the observations by taking time series from the simulation at $\pm 2R_E$ from Geotail's actual position to take this effect into account. Clearly, the simulation shows key observations, such as the earthward-tailward reversal of the flow and the sign reversal of B_z at substorm onset, the pressure and temperature peaks around onset, and the density dropout just after onset. The simulation results are often not more different from the observations than the differences in the measurements from the two plasma analyzers.

Figure 9b shows how the substorm evolution depends on parameters in the model. Specifically, we have varied the electron precipitation parameters (affect-

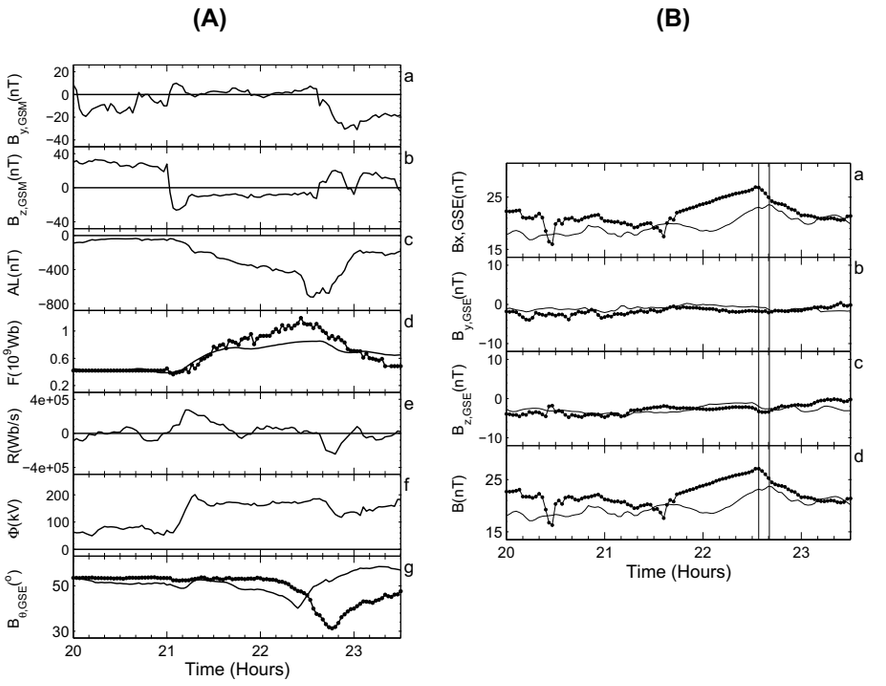


Fig. 8. **(A)** (a,b) Time series of the magnetic field B_y and B_z components in the magnetosheath at the subsolar magnetopause, (c) the modeled AL index, (d) the polar cap magnetic flux from the model (solid line) and from Polar Visible Imaging System (VIS) estimates (dotted line), (e) the rate of change of the polar cap flux as estimated from the model, (f) the cross polar cap potential from the model, and (g) the magnetic field elevation angle at GOES 8 from the model (solid line) and from GOES 8 (dotted line). **(B)** IMP 8 magnetic field data (dotted lines) and model results (solid lines) on November 24, 1996. GSE coordinates: (a) B_x, (b) B_y, (c) B_z, and (d) the total field. The two vertical lines mark the onset of the tail field reduction in the data and in the model, respectively

ing the ionospheric conductances) and parameters in the anomalous resistivity calculation. Clearly, the substorm evolution is very sensitive to these parameters. Only for specific combinations does a substorm develop. If the parameters are such that no substorm develops the magnetosphere enters into an enhanced convection mode during which very little energy is stored in the tail, but instead the energy is dissipated by enhanced reconnection in the mid tail. It is at this point not clear whether there is a universal set of parameters that would cause a substorm every time a substorm should occur. Thus, the parameter dependence is somewhat frustrating because it limits the model's predictive capabilities. On the flip side, however, it offers the opportunity to learn more about the substorm physics. Clearly, a certain amount of ionospheric field line tying is required to let a substorm occur. At the same time, intrinsic properties of the tail are also con-

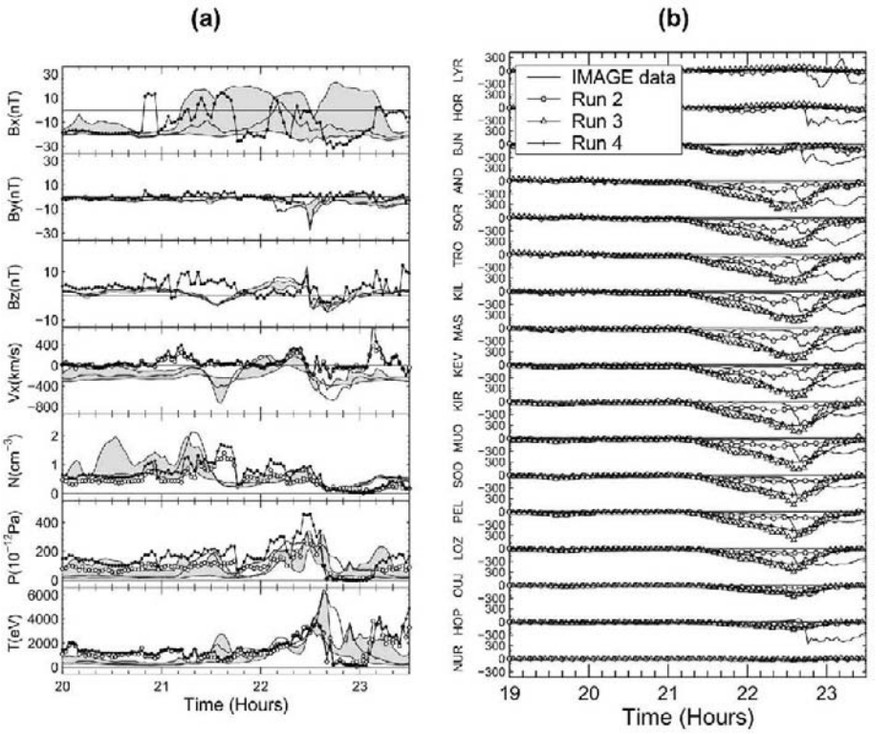


Fig. 9. **(a)** Geotail magnetic field and plasma data and comparison with the model results, on November 24, 1996. The dotted lines are magnetic field and plasma data from the Magnetic Field Experiment (MGF) and Low Energy Particle (LEP) experiment, respectively. Data represented with open circles are from the Comprehensive Plasma Instrumentation (CPI) instrument. Model results at the nominal Geotail position are drawn with a heavy solid line. The shaded area is bounded by time series taken from the model at locations that are 2 R_E above and below Geotail, respectively. **(b)** Comparison of IMAGE ground magnetometer recordings on November 24, 1996, with the results from three different simulation runs that did not produce a substorm (see text for details). The format of this figure is the same as that of Fig. 7b

trolling factors, in particular the onset of dissipation. Many more detailed case studies and theoretical investigation will be necessary to firmly establish these relationships. Note, that these results are not in contradiction with results from the recent "GEM reconnection challenge" [11,12,36,52,74,57], because we use a strongly nonlinear anomalous resistivity term and tail reconnection is indeed fast. Of course, the results from the "GEM reconnection challenge" offer new approaches to modeling the anomalous diffusion terms which will be pursued in the future.

The comparison of the results with data and the finding that the simulation captures the essence of the substorm development allows further investigation

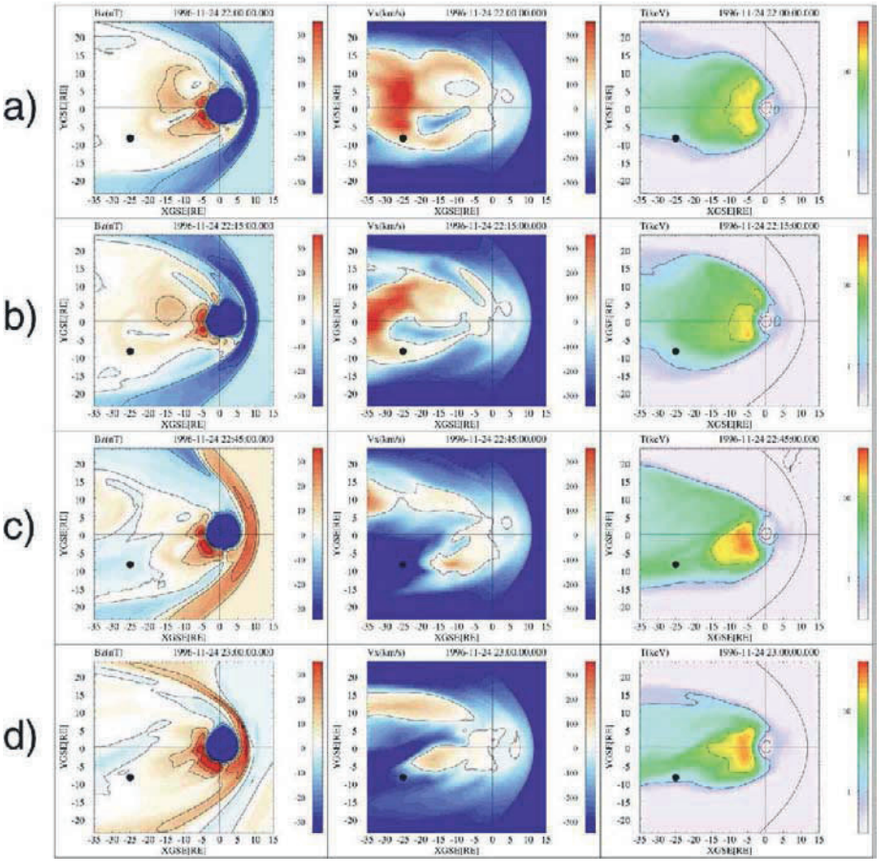


Fig. 10. The magnetic field B_z component, the velocity V_x component, and the plasma temperature in a plane at $Z_{\rm GSE}$=-3.3 R_E , at four different times: (a) 2200 UT, (b) 2215 UT, (c) 2245 UT, and (d) 2300 UT. Contours are drawn at the zero level for V_x, at 12 nT intervals for B_z, and at 1 and 10 keV for the temperature, respectively. The black dot marks Geotail's position.

of the processes in the tail that go beyond the view of *in situ* spacecraft observations. Figure 10 shows 2d cuts in a plane at $Z_{\rm GSE}$=-3.3 R_E which lies approximately in the center of the plasma sheet at 4 different times for the B_z component of the magnetic field, the V_x component of the velocity, and the plasma temperature. This figure demonstrates that the middle tail, where the energy conversion occurs, is highly structured. Unlike shown in many cartoons of tail reconnection, there is more than one x-line, the existing x-lines are not simply oriented in the y-direction, and the reconnection rate varies along any given x-line. Comparing B_z with V_x shows that channels of earthward flows transport flux earthward and lead to the dipolarization of the near-Earth field, however, in a strongly local time dependent manner. The simulation even cap-

tures the phenomenon of substorm particle injection. The plasma temperature near geosynchronous orbit rises significantly at substorm expansion onset. This shows that "particle injection" is, for the most part, earthward transport and adiabatic heating of plasma, because nothing more is included in the MHD description. Of course, kinetic features, like energy dispersion, are beyond the MHD description.

We should reiterate the fact that the value of this study lies in the combination of simulation with data analysis. The simulation itself would be hardly convincing, in particular, because it depends on the correct choice of parameters. On the other hand, the data itself are too sparse and allow for many different interpretations. Taken together, however, a clearer picture emerges that lends strong support to the near-Earth neutral line model of substorms with some modifications, such as fragmented x-lines and flow channels. This does not quite solve the "substorm problem" yet, because substorms come in many different flavors and because the simulation cannot yet capture all of the details, for example auroral arcs. However, with studies like these more progress can be made.

The simulation also shows why data analysis has not yet provided a clear picture (not even a penomenological one) of the substorm process and why intelligent people might come to quite different conclusions by looking at essentially the same data sets. Figures 9a and 10 show that even a small variation in Geotail's position leads to significantly different observations. One might be tempted to conclude that studies using observations from a single (or a few) spacecraft are doomed to fail to solve the problem because they can never derive suitable synoptic maps and at best murky statistics. Thus, a convincing solution may only be found with constellations of 10's to 100's of spacecraft, along with modeling and data assimilation [2,63]. More details about this substorm and the simulation can be found in [66,67].

4.2 Geomagnetic Storms

While substorms are manifestations of geomagnetic activity that is caused by relatively brief periods ($\sim$1h) of southward IMF, storms are caused by much longer (several hours to days) and stronger (negative IMF B_z of several 10's of nT) periods of southward IMF. In addition, the southward IMF of storms is often accompanied by high solar wind speed (sometimes over 1000 km s^{-1}) and high solar wind plasma density (several 10's cm^{-3}). Thus, storms exert much more stress on the magnetosphere and thus on any model as well. Storms are not only interesting because they cause extreme magnetospheric conditions, but also because they can cause severe space weather effects that are potentially harmful to our societal infrastructure (foremost geosynchronous and LEO satellites, power grids, and pipelines). Thus, the ability to forecast storm effects has recently become a topic of strong interest. Of course, global modeling in an operational setting would be one key element of space weather forecasting.

In the following we present a comparison of our model predictions with several key data sets for the Bastille Day storm (July 14/15, 2000), which was one of the

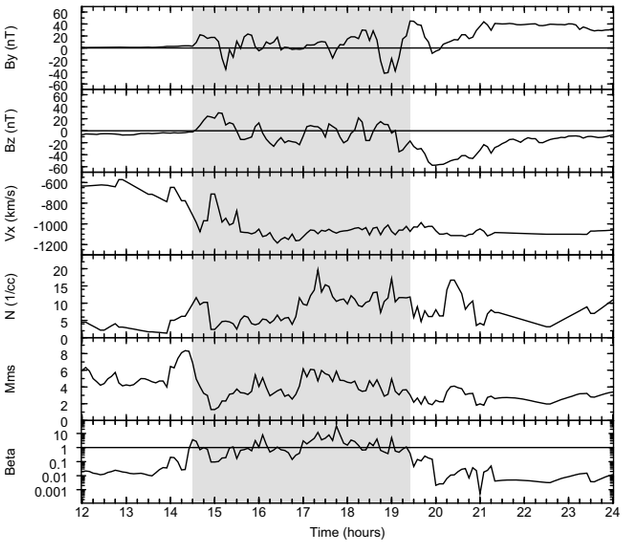


Fig. 11. Solar wind and IMF measurements for July 15, 2000. From top to bottom: The IMF B_y and B_z components, the solar wind velocity, the solar wind number density, the solar wind magnetosonic Mach number, and the plasma β (ratio of plasma pressure to magnetic field pressure). The shaded region is the CME sheath, which begins at the interplanetary shock and ends with the CME proper, which is defined here as the magnetic flux rope

strongest storms so far of this solar cycle [70]. We concentrate on space weather effects, that is, magnetospheric compression and ground magnetic perturbations.

Figure 11 shows the solar wind and IMF parameters for this event. The main characteristics are an interplanetary shock at ∼1430 UT, followed by the CME sheath until ∼1920 UT, followed by the CME proper, which lasts for almost 1 day. The solar wind speed reaches values of ∼1100 km s^{-1}, the density 20 cm^{-3}, and the IMF B_z −60 nT (at ∼2000 UT, the peak of the storm.)

Figure 12 shows the comparison of the magnetic field B_z component with measurements from three GOES geosynchronous satellites. First, compared to a typical quiet day (July 13, 2000), the field at these satellites is extremely distorted. In particular, the field measurements show several episodes in which B_z becomes negative, indicating that the spacecraft have crossed the magnetopause and entered the magnetosheath or even the solar wind. The simulation results compare extremely well with the observations, predicting all but a few of the magnetopause crossings.

Figure 13 shows the extreme compression and distortion of the magnetosphere at the peak of the storm. The magnetopause comes at this time as close as 4.9 R_E to Earth. At this time the dynamic pressure of the solar wind is 12 times as large as its average value, and the IMF B_z value is nearly −60 nT. The former causes a compression of the magnetosphere, while the latter causes

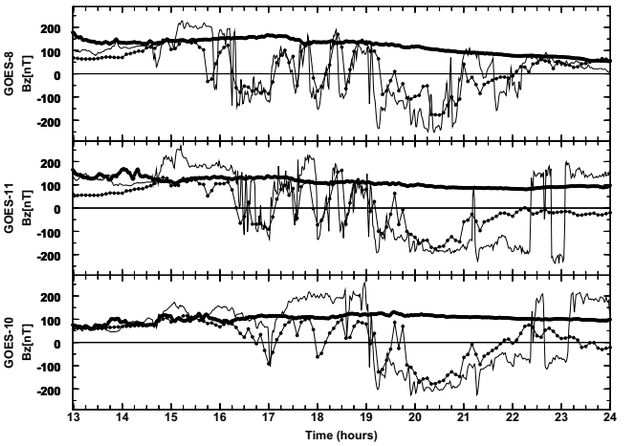


Fig. 12. The magnetic field component B_z observed by the 3 GOES satellites (from top to bottom: GOES 8, GOES 11, and GOES 10, ordered in increasing east longitude). The GOES data on July 15, 2000 are plotted as black solid lines, the corresponding model results as black lines with dots, and the GOES data for July 13, 2000 are plotted as thick black line. The latter data are from a typical geomagnetically quiet day

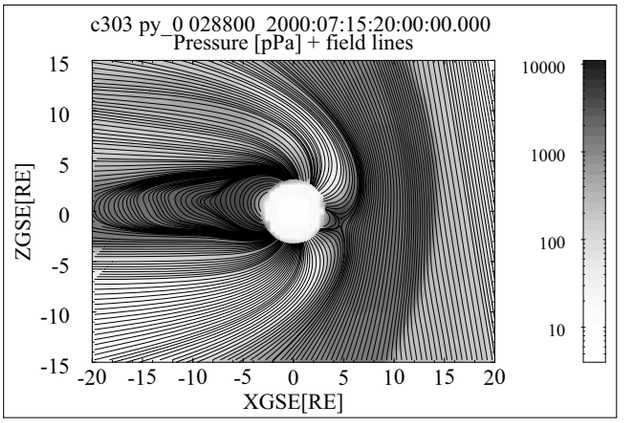


Fig. 13. Cut of the noon-midnight meridional plane shows the plasma pressure in grayscale and magnetic field lines at 2000 UT on July 15, 2000

flux erosion. Neither of these effects alone could bring the magnetopause this far in, but the combined effect does. Because storm effects of this magnitude are extremely rare (at most a few per solar cycle) events like this one put empirical models out of their valid parameter range, thus prediction is more reliable with numerical models.

The peculiar shape of the dayside magnetosphere has also consequences for the cross polar cap potential (CPCP). Simple prediction of the CPCP by re-

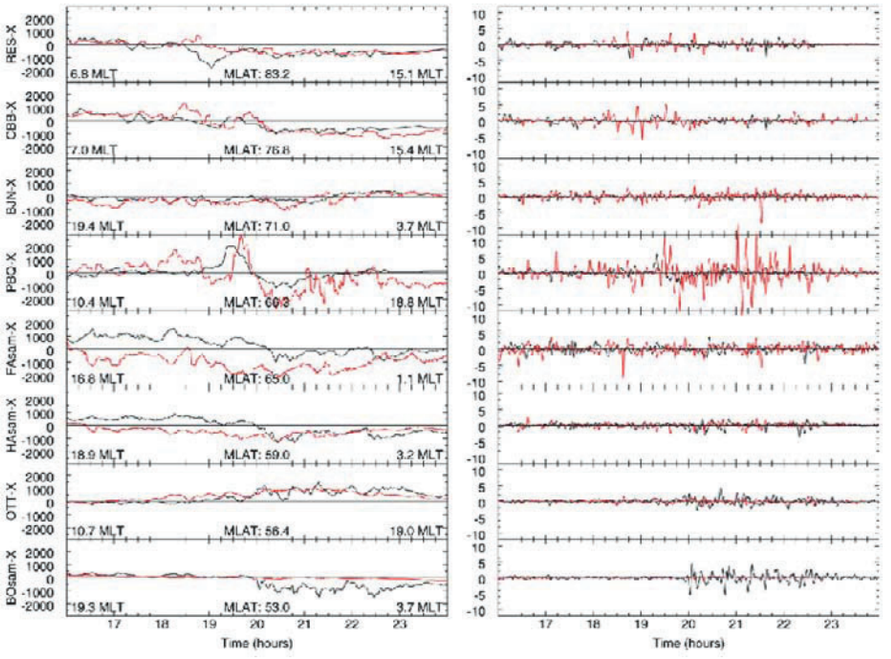


Fig. 14. The north-south ground magnetic perturbation (left column, the scale is in nT) and the corresponding time derivative (right column, the scale is in nT/s) for 8 ground magnetometer stations. The station names are indicated at the left of each panel, and the stations are ordered from north to south, although at different magnetic local times. Specifically, these stations are, with their abbreviation, geographic longitude and latitude given in parentheses: Resolute Bay (RES, 74.7°, 265.2°), Cambridge Bay (CBB, 69.1°, 255.0°), Bear Island (BJN, 74.52°, 19.02°), Poste-d.l.-Baleine (PBQ, 55.3°, 282.3°), Faroes (FAsam, 62.1°, 353.0°), Hankasalmi (HAsam, 62.3°, 26.7°), Ottawa (OTT, 45.4°, 284.5°), and Borok (BOsam, 58.0°, 38.3°). The black lines show the data and the red lines show the simulation results.

gression formulas [71,44] yield CPCP values of the order of 1000 kV. However, the model and measurements show that the CPCP reaches only ∼300 kV, thus, the CPCP saturates at this level. Figure 13 indicates indicates that because of the strong magnetopause erosion flux piles up in the lobes and the lobe shoulders bulge out. This causes the reconnection region to be partially shielded from the solar wind and the compressed IMF in the magnetosheath. Because of this shielding less magnetic energy can be transported to the reconnection site, leading to an overall reduction in the reconnection rate, and thus magnetospheric and ionospheric convection.

Figure 14 shows the comparison of ground magnetometer recordings with predictions from the simulation. The quality of the prediction varies significantly from station to station but is generally better at high and sub-auroral latitudes, and worst in the nightside auroral zone. Comparing the fluctuations, i.e., the time

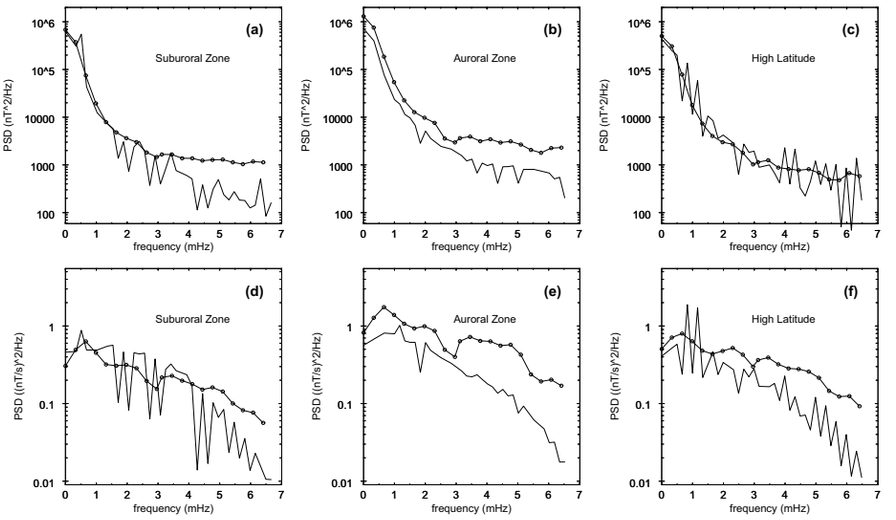


Fig. 15. (a-c) shows the averaged power spectral density (PSD, in units of $\mathrm{nT^2Hz^{-1}}$) of the north-south perturbation for ground magnetometer stations in the sub-auroral zone (a), the auroral zone (b), and the polar cap (c). The solid line is for the data, the solid line with dots represents the model result. The lower panel (d-f) shows the average PSD of the time derivative of the north-south perturbation (in units of $\mathrm{nT^2s^{-2}Hz^{-1}}$) for the same set of stations

derivative of the magnetic field (which are the cause of induced electric fields in power lines or pipelines) shows a similar picture. Figure 15 combines the spectral power of about 40 ground magnetometers in different latitude zones. Except for the auroral zone, the predictions of the power spectrum of the fluctuations is surprisingly good in the 0-3 mHz range. However, as said, this is not necessarily true for individual stations. In other words, the model predictions of the total power flux are more or less accurate; however, the model cannot yet predict the correct locations of large wave power flux. This event is discussed in more detail in [70].

4.3 Lessons from Data Comparisons

In the preceding sections we have shown that our model can reproduce a number of characteristic phenomena of substorms and storms. In particular, the model shows the magnetic energy loading in the tail during the substorm growth phase, the dipolarization of the near-Earth magnetic field in the expansion phase, the concurrent injection of energetic plasma into the inner magnetosphere, the formation of new x-lines, the formation and ejection of a plasmoid, and the intensification of the westward electrojet at expansion phase onset. Although these phenomena have all been observed in the past, no comprehensive (and universally accepted) model exists yet that puts these phenomena into context and

provides a satisfactory explanation. The fact that our model produces these phenomena, *and* the fact that these are all related to the onset of reconnection in the near-Earth tail at ∼20 R_E lends support to the Near Earth Neutral Line (NENL) model [53,6]. However, we also find that the tail dynamics is much more complex compared to the NENL prediction. In particular, the substorm onset is characterized by multiple x-lines that form at different local times and distances, and which produce rapid earthward and tailward plasma flows akin to Bursty Bulk Flows (BBFs) [1,4,3]. This fragmentation of the tail plasma sheet makes it all but impossible to solve the "2 minute problem," that is, the exact onset location and cause of the substorm expansion phase onset. We should also not forget that the model can not at present produce the brightening of the equator most auroral arc at onset, which is often cited as the first distinct substorm signature [47,46,27].

As far as magnetic storms are concerned, the study presented here should be viewed as a first step to model the magnetosphere under extreme solar wind conditions. This aspect is not only of particular importance for space weather, but also shows that the magnetosphere can exhibit a very particular behavior during storms. For example, both the simulation and the data show that the polar cap potential saturates when the driving solar wind ingredient, that is, the interplanetary electric field E_{ip}, reaches very high (>∼20 mV/m) values. For more benign conditions there is a more or less linear relationship between E_{ip} and the potential [71,15] which apparently loses its validity for the strong solar wind driving during storms. This result leads to a host of other questions, for example, if similar nonlinearities also exist for the energy input and dissipation in the magnetosphere. Global models face here a severe limitation because the most characteristic signature of a geomagnetic storm is the ring current which can not be treated adequately. Ring current formation is caused by the trapping of energetic plasma (>∼10 keV ions) in the inner part (2.5 – 8 R_E) of the magnetosphere by gradient and curvature drifts. The MHD formalism does not include these drifts; they are, in fact a consequence of non-Maxwellian distributions [34,35]. Thus, until global models are coupled with appropriate models of the inner magnetosphere (see below) such studies will be difficult.

5 Future Directions

The development of global geospace models has by no means come to an end. There are a large number of regions and processes that are not currently covered or that need improvement. At present some model developments appear most promising:

- The inclusion of the particle drift physics in the inner magnetosphere in the form of sub-models similar to the Rice Convection Model (RCM).
- Adaptive grid solvers that allow better resolution of plasma and field boundaries along with reduced numerical diffusion.
- Better magnetosphere-ionosphere coupling, in particular a more self-consistent model for electron precipitation.

- A multi-fluid formalism to study the ionospheric outflow and the role of ionospheric plasma in the magnetosphere. This includes the need for an outflow specification model or a self-consistent outflow model, none of which exists yet.
- Data assimilation from multiple spacecraft observations.

6 Conclusions

Global modeling has been proven an extremely powerful tool to study the solar-terrestrial plasma interaction. It's importance will likely increase in the future as models become ever more sophisticated and as computational power becomes ever more abundant and cheaper. We are now entering an era where global modeling does no longer depend on expensive supercomputers, but where affordable desktop equipment (Beowulf PC clusters) is becoming sufficient for meaningful global modeling. This should foster the more widespread use of global models, not only by the model developers themselves, but also by others in the scientific community.

Acknowledgements

This work was supported by grants ATM-0097143 and ATM-0084483 from the National Science Foundation and by grants NAG 5-10986 and NAG 5-12107 from the National Aeronautics and Space Administration.

References

1. V. Angelopoulos, W. Baumjohann, C. F. Kennel, F. V. Coroniti, M. G. Kivelson, R. Pellat, R. J. Walker, H. Lühr, and G. Paschmann. Bursty bulk flows in the inner plasma sheet. *J. Geophys. Res.*, 97:4027, 1992.
2. V. Angelopoulos, C. W. Carlson, D. W. Curtis, P. Harvey, R. P. Lin, F. S. Mozer, D. H. Pankow, J. Raeder, and C. T. Russell. On the necessity and feasability of a equatorial magnetospheric constellation. In V. Angelopoulos and P. V. Panetta, editors, *Science Closure and Enabling Technologies for Constellation Class Missions*, page 14. University of California, Berkeley, and NASA Goddard Space Flight Center, 1998.
3. V. Angelopoulos, C. F. Kennel, F. V. Coroniti, R. Pellat, M. G. Kivelson, R. J. Walker, C. T. Russell, W. Baumjohann, W. C. Feldman, and J. T. Gosling. Statistical chracteristics of bursty bulk flow events. *J. Geophys. Res.*, 99:21257, 1994.
4. V. Angelopoulos, Baumjohann W, C. F. Kennel, F. V. Coroniti, M. G. Kivelson, R. Pellat, R. J. Walker, H. Luhr, and G. Paschmann. Bursty bulk flows in the inner central plasma sheet. *J. Geophys. Res.*, 97:4027, 1992.
5. D. N. Baker, A. J. Klimas, D. Vassiliadis, T. I. Pulkkinen, and R. L. McPherron. Reexamination of driven and unloading aspects of magnetospheric substorms. *J. Geophys. Res.*, 102:7169, 1997.
6. D. N. Baker, T. I. Pulkkinen, V. Angelopoulos, W. Baumjohann, and R. L. McPherron. Neutral line model of substorms: Past results and present view. *J. Geophys. Res.*, 101:12975, 1996.

7. D. S. Balsara and D. Spicer. Maintaining pressure positivity in magnetohydrodynamic simulations. *J. Comp. Phys.*, 148:133, 1999.
8. A. R. Barakat and R. W. Schunk. Transport equations for multicomponent anisotropic space plasmas: A review. *Plasma Phys.*, 24:389, 1982.
9. M. J. Berger and P. Colella. Local adaptive mesh refinement for shock hydrodynamics. *J. Comp. Phys.*, 82:64, 1989.
10. M. J. Berger and J. Olinger. Adaptive mesh refinement for hyperbolic partial differential equations. *J. Comp. Phys.*, 53:484, 1984.
11. 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. Pritchett. Geospace Environmental Modeling (GEM) magnetic reconnection challenge. *J. Geophys. Res.*, 106:3715, 2001.
12. J. Birn and M. Hesse. Geospace Environmnent Modeling (GEM) magnetic reconnection challenge: resistive tearing, anisotropic pressure and hall effects. *J. Geophys. Res.*, 106:3737, 2001.
13. J. P. Boris. A physically motivated soluton of the Alfvèn problem, *nrl memorandum report 2167*, Naval Research Laboratory, Washington, D.C., 1970.
14. J. P. Boris and D. L. Book. Flux corrected transport: I. SHASTA, a fluid transport algorithm that works. *J. Comp. Phys.*, 11:38, 1973.
15. C. B. Boyle, P. H. Reiff, and M. R. Hairston. Empirical polar cap potentials. *J. Geophys. Res.*, 102:111, 1997.
16. J. U. Brackbill and D. C. Barnes. The effect of nonzero div B on the numerical solution of the magnetohydrodynamic equations. *J. Comp. Phys.*, 35:426, 1980.
17. S. H. Brecht. Global simulations using MHD codes: A few points to consider before you try one. *Space Sci. Rev.*, 42:169, 1985.
18. S. H. Brecht, J. G. Lyon, J. A. Fedder, and K. Hain. A time dependent three dimensional simulation of the Earth's magnetosphere: Reconnection events. *J. Geophys. Res.*, 87:6098, 1982.
19. M. Brio and C. C. Wu. An upwind differencing scheme for the equations of ideal magnetohydrodynamics. *J. Comp. Phys.*, 75:400, 1988.
20. W. Dai and P. R. Woodward. A high-order Godunov-type scheme for shock interactions in ideal magnetohydrodynamics. *SIAM J. Sci. Comp.*, 18:957, 1997.
21. C. R. DeVore. Flux-corrected transport techniques for multidimensional compressible magnetohydrodynamics. *Journal of Computational Physics*, 92:142, 1991.
22. C. R. Evans and J. F. Hawley. Simulation of magnetohydrodynamic flows: A constrained transport method. *Astrophys. J.*, 332:659, 1988.
23. J. A. Fedder and J. G. Lyon. The solar wind - magnetosphere - ionosphere current - voltage relationship. *Geophys Res. Lett.*, 14:880, 1987.
24. J. A. Fedder, S. P. Slinker, and J. G. Lyon. A comparison of global numerical simulation results to data for the January 27-28, 1992, geospace environment modeling challenge event. *J. Geophys. Res.*, 103:14799, 1998.
25. J. A. Fedder, S. P. Slinker, J. G. Lyon, and R. D. Elphinstone. Global numerical simulation of the growth phase and the expansion onset for a substorm observed by Viking. *J. Geophys. Res.*, 100:19083, 1995.
26. L. A. Frank, M. Ashour-Abdalla, J. Berchem, J. Raeder, W. R. Paterson, S. Kokubun, T. Yamamoto, R. P. Lepping, F. V. Coroniti, D. H. Fairfield, and K. L. Ackerson. Observations of plasmas and magnetic fields in Earth's distant magnetotail: Comparison with a global MHD model. *J. Geophys. Res.*, 100:19177, 1995.

27. L. A. Frank, W. R. Paterson, J. B. Sigwarth, and T. Mukai. Observations of plasma sheet dynamics earthward of the onset region with the Geotail spacecraft. *J. Geophys. Res.*, 106:18823, 2001.
28. T. J. Fuller-Rowell, D. Rees, S. Quegan, R. J. Moffett, M. V. Codrescu, and G. H. Millward. A coupled thermosphere-ionosphere model (CTIM). In R. W. Schunk, editor, *STEP Report*, page 217, NOAA/NGDC, Boulder, Colorado, 1996. Scientific Committee on Solar Terrestrial Physics (SCOSTEP).
29. T. I. Gombosi. *This book*, 2003.
30. T. I. Gombosi, K. G. Powell, and B. van Leer. Comment on "Modeling the magnetosphere for northward interplanetary magnetic field: Effects of numerical resistivity" by Joachim Raeder. *J. Geophys. Res.*, 105:13141, 2000.
31. A. Harten. On a class of high resolution total variation stable finite difference schemes. *SIAM J. Num. Anal.*, 21:1, 1984.
32. A. Harten, A. Hyman, and P. D. Lax. On finite difference approximations and entropy conditions for shocks. *Comm. Pure and Appl. Math.*, 29:297, 1976.
33. A. Harten and G. Zwas. Self-adjusting hybrid schemes for shock computations. *J. Comput. Phys.*, 9:568, 1972.
34. M. Heinemann. Role of collisionless heat flux in magnetospheric convection. *J. Geophys. Res.*, 104:28397, 1999.
35. M. Heinemann and R. A. Wolf. Relationships of models of the inner magnetosphere to the Rice Convection Model. *J. Geophys. Res.*, 106:15545, 2001.
36. M. Hesse, J. Birn, and M. Kuznetsova. Collisionless magnetic reconnection: electron process and transport modeling. *J. Geophys. Res.*, 106:3721, 2001.
37. C. Hirsch. *Numerical Computation of Internal and External Flow*, volume II. John Wiley, New York, 1990.
38. P. Janhunen, T. I. Pulkkinen, and K. Kauristie. Auroral fading in ionosphere-magnetosphere coupling model: Implications for possible mechanisms. *Geophys Res. Lett.*, 22:2049, 1995.
39. G.-S. Jiang and C. C. Wu. A high-order WENO finite difference scheme for the equations of idel magnetohydrodynamics. *J. Comp. Phys.*, 150:561, 1999.
40. S. Knight. Parallel electric fields. *Planet. Space Sci.*, 21:741, 1972.
41. P. D. Lax and B. Wendroff. Systems of conservation laws. *Comm. Pure Appl. Math.*, 13:217, 1960.
42. J. N. Leboeuf, T. Tajima, C. F. Kennel, and J. M. Dawson. Global simulation of the time dependent magnetosphere. *Geophys Res. Lett.*, 5:609, 1978.
43. X. Liu, S. Osher, and T. Chan. Weighted essentially non-oscillatory schemes. *J. Comp. Phys.*, 115:200, 1994.
44. G. Lu, P. H. Reiff, M. R. Hairston, R. A. Heelis, and J. L. Karty. Distribution of convection potential around the polar cap boundary as a function of the interplanetary magnetic field. *J. Geophys. Res.*, 94:13447, 1989.
45. A. T. Y. Lui. Extended consideration of a synthesis model for magnetic substorms. In J. R. Kan, T. A. Potemra, S. Kokubun, and T. Ijima, editors, *Magnetospheric Substorms*, volume 64, page 43. AGU Monogr. Ser., American Geophysical Union, 1991.
46. A. T. Y. Lui, R. E. Lopez, B. J. Anderson, K. Takahashi, L. J. Zanetti, R. W. McEntire, T. A. Potemra, D. M. Klumpar, greene, and strangeway. Current disruptions in the near-Earth neutral sheet region. *J. Geophys. Res.*, 97:1461, 1992.
47. A. T. Y. Lui, R. E. Lopez, S. M. Krimigis, R. W. McEntire, L. J. Zanetti, and T. A. Potemra. A case study of magnetotail current sheet disruption and diversion. *Geophys Res. Lett.*, 15:721, 1988.

48. J. G. Lyon, S. H. Brecht, J. A. Fedder, and P. J. Palmadesso. Computer simulation of a geomagnetic substorm. *Phys. Rev. Lett.*, 46:1038, 1981.
49. J. G. Lyon, R. E. Lopez, C. C. Goodrich, M. Wiltberger, and K. Papadopoulos. Simulation of the March 9, 1995, substorm: Auroral brightening and the onset of lobe reconnection. *Geophys Res. Lett.*, 25:3039, 1998.
50. L. R. Lyons. The Geospace Environment Modeling Grand Challenge. *J. Geophys. Res.*, 103:14781, 1998.
51. L. R. Lyons, R. L. McPherron, E. Zesta, G. D. Reeves, J. B. Sigwarth, and L.A. Frank. Timing of substorm signatures during the November 24, 1996 geospace environment modeling event. *J. Geophys. Res.*, submitted, 2000.
52. Z. W. Ma and A. Bhattacharjee. Hall magnetohydrodynamic reconnection: the geospace environment modeling challenge. *J. Geophys. Res.*, 106:3773, 2001.
53. R. L. McPherron. Physical processes producing magnetospheric substorms and magnetic storms. In J. Jacobs, editor, *Geomagnetism*, volume 4, page 593. Academic Press, 1991.
54. J. Moen and A. Brekke. The solar flux influence on quiet time conductances in the auroral ionosphere. *Geophys Res. Lett.*, 20:971, 1993.
55. T. Ogino. A three dimensional MHD simulation of the interaction of the solar wind with the Earth's magnetosphere: The generation of field aligned currents. *J. Geophys. Res.*, 91:6791, 1986.
56. T. Ogino, R. J. Walker, and M. Ashour-Abdalla. A global magnetohydrodynamic simulation of the response of of the magnetosphere to a northward turning of the interplanetary magnetic field. *J. Geophys. Res.*, 99:11027, 1994.
57. A. Otto. Geospace environment modeling (GEM) magnetic reconnection challenge: MHD and Hall MHD – constant and current dependent resistivity models. *J. Geophys. Res.*, 106:3751, 2001.
58. E. N. Parker. The alternative paradigm for magnetospheric physics. *J. Geophys. Res.*, 101:10587, 1996.
59. K. G. Powell, P. L. Roe, T. J. Linde, T. I. Gombosi, and D. L. DeZeeuw. A solution-adaptive upwind scheme for ideal magnetohydrodynamics. *J. Comp. Phys.*, 154:284, 1999.
60. J. Raeder. Global MHD simulations of the dynamics of the magnetosphere: Weak and strong solar wind forcing. In J. R. Kan, J. D. Craven, and S.-I. Akasofu, editors, *Proceedings of the Second International Conference on Substorms*, page 561. Geophysical Institute, Univ. of Alaska Fairbanks, 1995.
61. J. Raeder. Modeling the magnetosphere for northward interplanetary magnetic field: Effects of electrical resistivity. *J. Geophys. Res.*, 104:17357, 1999.
62. J Raeder. Reply. *J. Geophys. Res.*, 105:13149, 2000.
63. J. Raeder and V. Angelopoulos. Using global simulations of the magnetosphere for multi-satellite mission planning and analysis,. In V. Angelopoulos and P. V. Panetta, editors, *Science Closure and Enabling Technologies for Constellation Class Missions*, page 78. University of California, Berkeley, and NASA Goddard Space Flight Center, 1998.
64. J. Raeder, J. Berchem, and M. Ashour-Abdalla. The importance of small scale processes in global MHD simulations: Some numerical experiments. In T. Chang and J. R. Jasperse, editors, *The Physics of Space Plasmas*, volume 14, page 403, Cambridge, Mass., 1996. MIT Cent. for Theoret. Geo/Cosmo Plasma Phys.
65. J. Raeder, J. Berchem, M. Ashour-Abdalla, L. A. Frank, W. R. Paterson, K. L. Ackerson, R. P. Lepping, S. Kokubun, T. Yamamoto, and S. A. Slavin. Boundary layer formation in the magnetotail: Geotail observations and comparisons with a global MHD model. *Geophys Res. Lett.*, 24:951, 1997.

66. J. Raeder and N. Maynard. Foreword. *J. Geophys. Res.*, 106:345, 2001.
67. J. Raeder, R. L. McPherron, L. A. Frank, W. R. Paterson, J. B. Sigwarth, G. Lu, H. J. Singer, S. Kokubun, T. Mukai, and J. A. Slavin. Global simulation of the geospace environment modeling substorm challenge event. *J. Geophys. Res.*, 106:381, 2001.
68. J. Raeder, R. J. Walker, and M. Ashour-Abdalla. The structure of the distant geomagnetic tail during long periods of northward IMF. *Geophys Res. Lett.*, 22:349, 1995.
69. J. Raeder, Y. Wang, and T. Fuller-Rowell. Geomagnetic storm simulation with a coupled magnetosphere - ionosphere - thermosphere model. In P. Song, G. Siscoe, and H. J. Singer, editors, *Space Weather, AGU Geophys. Monogr. Ser.*, volume 125, page 377. American Geophysical Union, 2001.
70. J. Raeder, Y. L. Wang, T. J. Fuller-Rowell, and H. J. Singer. Global simulation of space weather effects of the Bastille Day storm. *Sol. Phys.*, 204:325, 2001.
71. P. H. Reiff, R. W. Spiro, and T. W. Hill. Dependence of polar cap potential drop of interplanetary parameters. *J. Geophys. Res.*, 86:7639, 1981.
72. A. D. Richmond and R. G. Roble. Electrodynamic effects of thermospheric winds for the NCAR thermospheric general circulation model. *J. Geophys. Res.*, 92:12365, 1987.
73. R. M. Robinson, R. R. Vondrak, K. Miller, T. Dabbs, and D. Hardy. On calculating ionospheric conductances from the flux and energy of precipitating electrons. *J. Geophys. Res.*, 92:2565, 1987.
74. M. A. Shay, J. F. Drake, B. N. Rogers, and R. E. Denton. Alfvenic collisionless magnetic reconnection and the Hall term. *J. Geophys. Res.*, 106:3759, 2001.
75. G. A. Sod. *Numerical Methods in Fluid Dynamics.* Cambridge University Press, Cambridge, 1985.
76. B. U. O. Sonnerup and L. J. Cahill. Magnetopause structure and attitude from Explorer 12 observations. *J. Geophys. Res.*, 72:171, 1967.
77. B. U. O. Sonnerup and L. J. Cahill. Explorer 12 observations of the magnetopause current layer. *J. Geophys. Res.*, 73:1757, 1968.
78. R. J. Strangeway, R. C. Elphic, W. J. Peria, and C. W. Carlson. FAST observations of electromagnetic stresses applied to the polar ionosphere. volume 118, page 21. AGU Monogr. Ser., American Geophysical Union, 1999.
79. R. J. Strangeway and J. Raeder. On the transition from collisionless to collisional magnetohydrodynamics. *J. Geophys. Res.*, 106:1955, 2001.
80. P. K. Sweby. High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM J. Num. Anal.*, 21:995, 1984.
81. T. Tanaka. Generation mechanisms for magnetosphere-ionosphere current systems deduced from a three-dimensional MHD simulation of the solar wind-magnetosphere-ionosphere coupling processes. *J. Geophys. Res.*, 100:12057, 1995.
82. G. Toth. The $\nabla \cdot B$ constraint in shock-capturing magnetohydrodynamics codes. *J. Comp. Phys.*, 161:605, 2000.
83. A. Usadi, A. Kageyama, K. Watanabe, and T. Sato. A global simulation of the magnetosphere with a long tail: Southward and northward interplantary magnetic field. *J. Geophys. Res.*, 98:7503, 1993.
84. B. Van Leer. Towards the ultimate conservative difference scheme. I. The quest for monotonicity. In *Lecture Notes in Physics*, volume 18, page 163. Springer Verlag, Berlin, 1973.
85. B. Van Leer. Towards the ultimate conservative difference scheme. II. Monotonicity and conservation combined in a second order scheme. *J. Comp. Phys.*, 14:361, 1974.

86. B. Van Leer. Towards the ultimate conservative difference scheme. III. Upstream centered finite difference schemes for ideal compressible flow. *J. Comp. Phys.*, 23:263, 1977.
87. V. M. Vasyliunas. Mathematical models of magnetospheric convection and its coupling to the ionosphere. In *Particles and Fields in the Magnetosphere*, page 61. Dordrecht, Netherlands: D. Reidel, 1970.
88. K. Watanabe and T. Sato. Global simulation of the solar wind-magnetosphere interaction: The importance of its numerical validity. *J. Geophys. Res.*, 95:75, 1990.
89. R. A. Wolf. The quasi-static (slow flow) region of the magnetosphere. In R. L. Carovillano and J. M. Forbes, editors, *Solar Terrestrial Physics*, page 303. D. Reidel, Hingham, MA, 1983.
90. H. C. Yee. On symmetric and upwind TVD schemes. In *Proc. 6th GAMM Conference on Numerical Methods in Fluid Dynamics*, page 399, Braunschweig, 1985. Vieweg.
91. H. C. Yee. Construction of explicit and implicit symmetric TVD schemes and their applications. *J. Comput. Phys.*, 68:151, 1987.
92. A. L. Zachary, A Malagoli, and P. Colella. A higher-order Godunov method for multi-dimensional ideal magnetohydrodynamics. *SIAM J. Sci. Comp.*, 15:263, 1994.
93. S. T. Zalesak. Fully multidimensional flux-corrected transport. *J. Comp. Phys.*, 31:355, 1979.
94. S. T. Zalesak. Very high order pseudospectral flux-corrected transport (FCT) algorithms for conservation laws. In R. Vichnevetsky and R. S. Stepleman, editors, *Proceedings of the Fourth IMACS International Symposium on Computer Methods for Partial Differential Equations*, page 126, New Brunswick, 1981. IMACS, Rutgers University.

Adaptive Mesh Refinement for Global Magnetohydrodynamic Simulation

Tamas I. Gombosi, Darren L. De Zeeuw, Kenneth G. Powell, Aaron J. Ridley, Igor V. Sokolov, Quentin F. Stout, and Gábor Tóth*

Center for Space Environment Modeling, The University of Michigan, Ann Arbor, Michigan, USA

Abstract. The first part of this paper reviews some physics issues representing major computational challenges for global MHD models of the space environment. These issues include: (i) mathematical formulation and discretization of the governing equations that ensure the proper jump conditions and propagation speeds, (ii) regions of relativistic Alfvén speed, (iii) regions dominated by strong intrinsic planetary magnetic field with strong gradients, and (iv) the religiously debated issue of controling the divergence of the magnetic field. The second part of the paper concentrates to modern solution methods that have been developed by the aerodynamics, applied mathematics and DoE communities. Such methods have recently begun to be implemented in space-physics codes, which solve the governing equations for a compressible magnetized plasma. These techniques include high-resolution upwind schemes, block-based solution-adaptive grids and domain decomposition for parallelization. While some of these techniques carry over relatively straightforwardly to space physics, space physics simulations pose some new challenges. We give a brief review of the state-of-the-art in modern space-physics codes. Finally, we describe the space physics MHD code developed at the University of Michigan and its recent coupling to a thermosphere-ionosphere and inner magnetosphere model.

1 Introduction

Global computational models based on first principles represent a very important component of efforts to understand the intricate processes coupling the Sun to the geospace environment. The hope for such models is that they will eventually fill the gaps left by measurements, extending the spatially and temporarily limited observational database into a self-consistent global understanding of our space environment.

Presently, and in the foreseeable future, magnetohydrodynamic (MHD) models are the only models that can span the enormous distances present in the magnetosphere. However, it should not be forgotten that even generalized MHD equations are only a relatively low-order approximation to more complete physics; they provide only a simplified description of natural phenomena in space plasmas.

* also at Department of Atomic Physics, Loránd Eötvös University, Budapest, Hungary

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 247–274, 2003.

2 Non-relativistic Magnetohydrodynamics

The governing equations for an ideal, non-relativistic, compressible plasma may be written in a number of different forms. While the different forms of the MHD equations describe the same physics at the differential equation level, there are important practical differences when one solves discretized forms of the various formulations.

According to the Lax-Wendroff theorem [37] only conservative schemes can be expected to get the correct jump conditions and propagation speed for a discontinuous solution. This fact is much less emphasized in the global magnetosphere simulation literature than the more controversial divergence of B issue. In some test problems the non-conservative discretization of the MHD equations can lead to significant errors, which do not diminish with increased grid resolution.

2.1 Primitive Variable Form

In primitive variables, the governing equations of ideal magnetohydrodynamics, which represent a combination of the Euler equations of gasdynamics and the Maxwell equations of electromagnetics, may be written as:

$$\frac{\partial \rho}{\partial t} + \mathbf{u} \cdot \nabla \rho + \rho \nabla \cdot \mathbf{u} = 0 \tag{1a}$$

$$\rho \frac{\partial \mathbf{u}}{\partial t} + \rho \mathbf{u} \cdot \nabla \mathbf{u} + \nabla p - \frac{1}{\mu_0} \mathbf{j} \times \mathbf{B} = 0 \tag{1b}$$

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} = 0 \tag{1c}$$

$$\frac{\partial p}{\partial t} + \mathbf{u} \cdot \nabla p + \gamma p \nabla \cdot \mathbf{u} = 0 \tag{1d}$$

where μ_0 and γ represent the magnetic permeability of vacuum and the specific heat ratio of the gas. In addition, the current density, $\mathbf{j}$, and the electric field vector, $\mathbf{E}$, are related to the magnetic field $\mathbf{B}$ by *Ampère's law* and *Ohm's law*:

$$\mathbf{j} = \frac{1}{\mu_0} \nabla \times \mathbf{B} \tag{2a}$$

$$\mathbf{E} = -\mathbf{u} \times \mathbf{B} \tag{2b}$$

2.2 Gasdynamic Conservation Form

For one popular class of schemes, the equations are written in a form in which the gasdynamic terms are put in divergence form, and the electromagnetic terms in the momentum and energy equations are treated as source terms. This gives:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \tag{3a}$$

$$\frac{\partial (\rho \mathbf{u})}{\partial t} + \nabla \cdot (\rho \mathbf{u}\mathbf{u} + p\mathbf{I}) = \frac{1}{\mu_0} \mathbf{j} \times \mathbf{B} \tag{3b}$$

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} = 0 \tag{3c}$$

$$\frac{\partial E_{gd}}{\partial t} + \nabla \cdot [\mathbf{u} (E_{gd} + p)] = \frac{1}{\mu_0} \mathbf{j} \cdot \mathbf{E} \tag{3d}$$

where $\mathbf{I}$ is the identity matrix and E_{gd} is the gasdynamic total energy, given by

$$E_{gd} = \frac{1}{2}\rho\, u^2 + \frac{1}{\gamma - 1}\, p \tag{4}$$

This formulation is used in several popular magnetosphere codes [20,49], as well as in the publicly available Zeus code [63].

2.3 Fully Conservative Form

The fully conservative form of the equations is

$$\frac{\partial \mathbf{U}}{\partial t} + (\nabla \cdot \mathbf{F})^{\mathrm{T}} = 0\ , \tag{5}$$

where $\mathbf{U}$ is the vector of conserved quantities and $\mathbf{F}$ is a flux diad,

$$\mathbf{U} = \begin{pmatrix} \rho \\ \rho\mathbf{u} \\ B \\ E_{mhd} \end{pmatrix} \tag{6a}$$

$$\mathbf{F} = \begin{pmatrix} \rho\mathbf{u} \\ \rho\mathbf{u}\mathbf{u} + \left(p + \frac{1}{2\mu_0}B^2\right)\mathbf{I} - \frac{1}{\mu_0}\mathbf{B}\mathbf{B} \\ \mathbf{u}\mathbf{B} - \mathbf{B}\mathbf{u} \\ \mathbf{u}\left(E_{mhd} + p + \frac{1}{2\mu_0}B^2\right) - \frac{1}{\mu_0}\left(\mathbf{u}\cdot\mathbf{B}\right)\mathbf{B} \end{pmatrix}^{\mathrm{T}} \tag{6b}$$

where E_{mhd} is the magnetohydrodynamic energy, given by

$$E_{mhd} = E_{gd} + \frac{1}{2\mu_0}B^2 \tag{7}$$

2.4 Symmetrizable Formulation

Symmetrizable systems of conservation laws have been studied by Godunov [21] and Harten [30], among others. One property of the symmetrizable form of a system of conservation laws is that an added conservation law

$$\frac{\partial\left(\rho s\right)}{\partial t} + \frac{\partial\left(\rho\, s\, u_x\right)}{\partial x} + \frac{\partial\left(\rho\, s\, u_y\right)}{\partial y} + \frac{\partial\left(\rho\, s\, u_z\right)}{\partial z} = 0$$

for the entropy s can be derived by a linear combintion of the system of equations. For the ideal MHD equations, as for the gasdynamic equations, the entropy is $s = \log(p/\rho^{\gamma})$. Another property is that the system is Galilean invariant; all waves in the system propagate at speeds $u \pm c_w$ (for MHD, the possible values of c_w are the Alfvén, magnetofast and magentoslow speeds). Neither of these properties holds for the fully conservative form of the MHD equations.

Godunov showed that the fully conservative form of the MHD equations (eq. 5) is not symmetrizable [21]. The symmetrizable form may be written as

$$\frac{\partial \mathbf{U}}{\partial t} + (\nabla \cdot \mathbf{F})^{\mathrm{T}} = \mathbf{Q} \,, \tag{8}$$

where

$$\mathbf{Q} = -\nabla \cdot \mathbf{B} \begin{pmatrix} 0 \\ \frac{1}{\mu_0}\mathbf{B} \\ \mathbf{u} \\ \frac{1}{\mu_0}\mathbf{u} \cdot \mathbf{B} \end{pmatrix} \tag{9}$$

Vinokur separately showed that eq. (8) can be derived starting from the primitive form, if no stipulation is made about $\nabla \cdot \mathbf{B}$ in the derivation. Powell showed that this symmetrizable form can be used to derive a Roe-type approximate Riemann solver for solving the MHD equations in multiple dimensions [45].

The MHD eigensystem arising from eq. (5) or eq. (8) leads to eight eigenvalue/eigenvector pairs. The eigenvalues and associated eigenvectors correspond to an entropy wave, two Alfvén waves, two magnetofast waves, two magnetoslow waves, and an eighth eigenvalue/eigenvector pair that depends on which form of the equations is being solved. This last wave (which describes the jump in the normal component of the magnetic field at discontinuities) has a zero eigenvalue in the fully conservative case, and an eigenvalue equal to the normal component of the velocity, u_n, in the symmetrizable case. The expressions for the eigenvectors, and the scaling of the eigenvectors, are more intricate than in gasdynamics [59].

We note that while eq.(5) is fully conservative, the symmetrizable formulation (given by eq. 8) is formally not fully conservative. Terms of order $\nabla \cdot \mathbf{B}$ are added to what would otherwise be a divergence form. The danger of this is that shock jump conditions may not be correctly met, unless the added terms are small, and/or they alternate in sign in such a way that the errors are local, and in a global sense cancel in some way with neighboring terms. This downside, however, has to be weighed against the alternative; a system (i.e., the one without the source term) that, while conservative, is not Gallilean invariant, has a zero eigenvalue in the Jacobian matrix, and is not symmetrizable.

3 Semi-relativistic Plasmas

While the solar-wind speed remains non-relativistic in the solar system, the intrinsic magnetic fields of several planets in the solar system are high enough, and the density of the solar wind low enough, that the Alfvén speed,

$$V_{\mathrm{A}} = \sqrt{\frac{B^2}{\mu_0 \rho}} \tag{10}$$

can reach appreciable fractions of the speed of light. In the case of Jupiter, the Alfvén speed in the vicinity of the poles is of order ten! Even Earth has a strong

enough intrinsic magnetic field that the Alfvén speed reaches twice the speed of light in Earth's near-auroral regions.

3.1 Limiting the Alfvén Speed

For these regions, solving the non-relativistic ideal MHD equations does not make sense. Having waves in the system propagating faster than the speed of light, besides being non-physical, causes a number of numerical difficulties. However, solving the fully relativistic MHD equations is overkill. What is called for is a semi-relativistic form of the equations, in which the flow speed and acoustic speed are non-relativistic, but the Alfvén speed can be relativistic. A derivation of these semi-relativistic equations from the fully relativistic equations is given in [25]; the final result is presented here.

The semi-relativistic ideal MHD equations are of the form

$$\frac{\partial \mathbf{U}_{sr}}{\partial t} + (\nabla \cdot \mathbf{F}_{sr})^{\mathrm{T}} = 0 \tag{11}$$

where the state vector, $\mathbf{U}_{sr}$, and the flux diad, $\mathbf{F}_{sr}$, are

$$\mathbf{U}_{sr} = \begin{pmatrix} \rho \\ \rho\mathbf{u} + \frac{1}{c^2}\mathbf{S}_{\mathrm{A}} \\ \mathbf{B} \\ \frac{1}{2}\rho u^2 + \frac{1}{\gamma-1}p + e_{\mathrm{A}} \end{pmatrix} \tag{12a}$$

$$\mathbf{F}_{sr} = \begin{pmatrix} \rho\mathbf{u} \\ \rho\mathbf{u}\mathbf{u} + p\mathbf{I} + \mathbf{P}_{\mathrm{A}} \\ \mathbf{u}\,\mathbf{B} - \mathbf{B}\,\mathbf{u} \\ \left(\frac{1}{2}\rho u^2 + \frac{\gamma}{\gamma-1}p\right)\mathbf{u} + \mathbf{S}_{\mathrm{A}} \end{pmatrix}^{\mathrm{T}} \tag{12b}$$

In the above,

$$\mathbf{S}_{\mathrm{A}} = \frac{1}{\mu_0}\left(\mathbf{E} \times \mathbf{B}\right) \tag{13a}$$

$$e_{\mathrm{A}} = \frac{1}{2\mu_0}\left(B^2 + \frac{1}{c^2}E^2\right) \tag{13b}$$

$$\mathbf{P}_{\mathrm{A}} = e_{\mathrm{A}}\mathbf{I} - \frac{1}{\mu_0}\mathbf{B}\,\mathbf{B} - \frac{1}{\mu_0 c^2}\mathbf{E}\,\mathbf{E} \tag{13c}$$

are the Poynting vector, the electromagnetic energy density, and the electromagnetic pressure tensor, respectively. The electric field $\mathbf{E}$ is related to the magnetic field $\mathbf{B}$ by Ohm's law, (eq. 2b).

3.2 Lowering the Speed of Light

This new system of equations has wave speeds that are limited by the speed of light; for strong magnetic fields, the modified Alfvén speed (and the modified magnetofast speed) asymptote to c. The modified magnetoslow speed asymptotes

to a, the acoustic speed. This property offers the possibility of a rather tricky convergence-acceleration technique for explicit time-stepping schemes, first suggested by Boris [11]; the wave speeds can be lowered, and the stable time-step thereby raised, by artificially lowering the value taken for the speed of light. This method is known as the "Boris correction."

The equations in Sect. 3.1 are valid in physical situations in which $V_A > c$. A slight modification yields a set of equations, the steady-state solutions of which are independent of the value taken for the speed of light. Defining the true value of the speed of light to be c_0, to distinguish it from the artificially lowered speed of light, c, the equations are:

$$\frac{\partial \mathbf{U}_{sr}}{\partial t} + (\nabla \cdot \mathbf{F}_{sr})^{\mathrm{T}} = \mathbf{Q}_{c_0} \tag{14}$$

where the state vector, $\mathbf{U}_{sr}$, and the flux diad, $\mathbf{F}_{sr}$, are as defined above, and the new source term in the momentum equation is

$$\mathbf{Q}_{c_0} = \frac{1}{\mu_0}\left(\frac{1}{c_0^2} - \frac{1}{c^2}\right)\mathbf{E}\nabla \cdot \mathbf{E} \tag{15}$$

An implementation of the semi-relativistic equations has been made in the BATSRUS code developed at the University of Michigan [46,25].

4 Splitting the Magnetic Field

For problems in which strong externally imposed magnetic fields are present, accuracy can be increased by solving for the deviation of the magnetic field from this prescribed component. For instance, in magnetosphere-type simulations a strong dipole-like magnetic field dominates the solution near the body. Solving for the deviation $\mathbf{B}_1$ from the embedded field $\mathbf{B}_0$ is inherently more accurate than solving for the full magnetic field vector $\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_1$. This approach was first suggested by Ogino and Walker [42], applied to Godunov-type schemes by Tanaka [64] and later employed by our group [46]. Recently we generalized it to semi-relativistic MHD [25].

The full magnetic field vector $\mathbf{B}$ can be written as

$$\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_1 \tag{16}$$

where $\mathbf{B}_0$ is given analytically and thus $\nabla \cdot \mathbf{B}_0 = 0$, while $\mathbf{B}_1$ is calculated by the numerical scheme. Note that $\mathbf{B}_1$ is not necessarily small relative to $\mathbf{B}_0$. We also introduce the non-relativistic current density $\mathbf{j}_0 = (1/\mu_0)\nabla \times \mathbf{B}_0$. The splitting is most important when the equations are solved in a (near) conservation form, since the total energy density can be completely dominated by the magnetic energy $\mathbf{B}_0^2/(2\mu_0)$. When the pressure is calculated from the total energy density, it can easily become negative, as we take differences of two huge numbers to obtain a small one. This problem can be mitigated by rewriting the energy equation in terms of the modified total energy density

$$e_1 = E_{gd} + \frac{1}{2\mu_0}\left(B_1^2 + \frac{1}{c^2}E^2\right) \tag{17}$$

Note that the electric energy still contains contribution from $\mathbf{B}_0$, but that is reduced by the factor $(1/c^2)$. With these definitions the conservation form of the semi-relativistic MHD equations (eq. 11) can be rewritten as

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \tag{18a}$$

$$\frac{\partial}{\partial t}\left(\rho \mathbf{u} + \tfrac{1}{c^2}\mathbf{S}_{\mathrm{A}}\right) + \nabla \cdot [\rho \mathbf{u}\mathbf{u} + \mathbf{I}p + \mathbf{P}_{\mathrm{A},1}] = \mathbf{j}_0 \times \mathbf{B}_0 \tag{18b}$$

$$\frac{\partial \mathbf{B}_1}{\partial t} + \nabla \cdot [\mathbf{u}\mathbf{B} - \mathbf{B}\mathbf{u}] = -\frac{\partial \mathbf{B}_0}{\partial t} \tag{18c}$$

$$\frac{\partial e_1}{\partial t} + \nabla \cdot \left[\mathbf{u}\left(\tfrac{1}{2}\rho \mathbf{u}^2 + \frac{\gamma p}{\gamma - 1}\right) + \tfrac{1}{\mu_0}\mathbf{E} \times \mathbf{B}_1\right] = -\mathbf{B}_1 \cdot \frac{\partial \mathbf{B}_0}{\partial t} + \mathbf{E} \cdot \mathbf{j}_0 \tag{18d}$$

where

$$P_{\mathrm{A},1} = \frac{1}{\mu_0}\left[\mathbf{I}\left(\frac{1}{2}\mathbf{B}_1^2 + \mathbf{B}_1 \cdot \mathbf{B}_0 + \frac{1}{2c^2}\mathbf{E}^2\right) - \mathbf{B}_1\mathbf{B}_1 - \mathbf{B}_1\mathbf{B}_0 - \mathbf{B}_0\mathbf{B}_1\right] \tag{19}$$

The splitting did not modify the continuity equation. In the momentum equation the dominant $B_0^2\mathbf{I}$ and $\mathbf{B}_0\mathbf{B}_0$ terms are moved into the source term $\mathbf{j}_0 \times \mathbf{B}_0$, which can be calculated analytically, and it is identically zero if $\mathbf{B}_0$ is a force free field. The induction equation is modified in a trivial way, by moving the time derivative of $\mathbf{B}_0$ to the right hand side. Again, this term can be calculated analytically, and in the case of a stationary $\mathbf{B}_0$ field, it vanishes. The split energy equation is obtained after quite some algebra. Most of the dominant B_0^2 and $\mathbf{E} \times \mathbf{B}_0$ terms are eliminated, but the remaining source terms $\mathbf{B}_1 \cdot \partial \mathbf{B}_0/\partial t$ and $\mathbf{E} \cdot \mathbf{j}_0$ contain the numerically calculated $\mathbf{B}_1$ and $\mathbf{u}$ quantities. In case of a potential ($\mathbf{j}_0 = 0$) and/or stationary $\mathbf{B}_0$ field one or both energy source terms can be eliminated.

One may add the source terms involving $\nabla \cdot \mathbf{B}$ to the split momentum, induction and/or energy equations (18b)−(18d) if the numerical scheme does not keep $\nabla \cdot \mathbf{B}$ exactly zero. Of course, $\nabla \cdot \mathbf{B} = \nabla \cdot \mathbf{B}_1$, since the analytic $\mathbf{B}_0$ field must be divergence free by definition.

5 Solution Techniques

5.1 Finite-Volume Scheme

The MHD equations are well suited for finite volume methods when the governing equations are integrated over a computational cell i, yielding

$$\frac{d\mathbf{U}_i}{dt} = -\frac{1}{V_i}\sum_{\text{faces}} \mathbf{F} \cdot \hat{\mathbf{n}} A - \frac{\mathbf{Q}_i}{V_i}\sum_{\text{faces}} \mathbf{B} \cdot \hat{\mathbf{n}} A\,, \tag{20}$$

where V_i is the volume of cell i, A is the surface area of the faces forming the computational cell, $\hat{\mathbf{n}}$ is the unit vector normal to the cell faces, $\mathbf{U}_i$ is the cell-averaged conserved solution vector, and $\mathbf{Q}_i$ is given by

$$\mathbf{Q}_i = -\begin{bmatrix} 0 \\ \frac{1}{\mu_0}\mathbf{B}_i \\ \mathbf{u}_i \\ \frac{1}{\mu_0}\mathbf{u}_i \cdot \mathbf{B}_i \end{bmatrix}. \tag{21}$$

The numerical face fluxes, $\mathbf{F}\cdot\hat{\mathbf{n}}$, are defined in terms of the left and right interface solution states, $\mathbf{U}_{\mathrm{L}}$ and $\mathbf{U}_{\mathrm{R}}$, as follows

$$\mathbf{F}\cdot\hat{\mathbf{n}} = \mathcal{F}\left(\mathbf{U}_{\mathrm{L}}, \mathbf{U}_{\mathrm{R}}, \hat{\mathbf{n}}\right) , \tag{22}$$

where $\mathbf{U}_{\mathrm{L}}$ and $\mathbf{U}_{\mathrm{R}}$ are the state vectors at the left and right sides of the interface.

5.2 TVD-MUSCL

Because the MHD equations are a system of hyperbolic conservation laws, many of the techniques that have been developed for the Euler equations can be applied relatively straightforwardly. In particular, the high-resolution finite-volume approach of van Leer [69] (i.e. approximate Riemann solver + limited interpolation scheme + multi-stage time-stepping scheme) is perfectly valid. The Rusanov/Lax-Friedrichs approximate Riemann solver can be applied directly; no knowledge of the eigensystem of the MHD equations is required other than the fastest wave speed in the system. A Roe-type scheme [58] can be constructed for non-relativistic MHD, but requires more work, because of the complexity of the eigensystem. In addition, an HLLE-type Riemann solver has been derived by Linde [39]; it is less dissipative than the Rusanov/Lax-Friedrichs scheme, but less computationally intensive than the Roe scheme. Whichever approximate Riemann solver is chosen to serve as the flux function, standard interpolation schemes and limiters can be used to construct a finite-volume scheme.

5.3 Pressure Positivity

One added difficulty in solving the MHD equations is that the MHD energy has three components: internal, magnetic and kinetic. Thus, as in gasdynamics, flows with substantially more kinetic energy than internal energy can lead to positivity problems when computing the pressure. Also, in contrast to gasdynamics, regions in which the magnetic field is large can yield similar problems. Conservative and positive HLLE-type schemes for MHD have been described by Janhunen [32]. Another alternative, due to Balsara and Spicer [2], is to use a hybrid scheme: both the conservative energy equation and the entropy equations are solved. Close to shock waves the energy equation is used to obtain the correct weak solution, at other places the more robust and positive entropy equation can be used. A variant of this technique has been implemented in our code.

6 Controlling $\nabla\cdot\mathbf{B}$

Another way in which the numerical solution of the MHD equations differs from that of the gasdynamic equations is the constraint that $\nabla\cdot\mathbf{B} = 0$. Enforcing this constraint numerically, particularly in shock-capturing codes, can be done in a number of ways, but each way has its particular strengths and weaknesses. Only a brief overview is given below; each of the schemes discussed below is explained more fully in the references cited, and Tóth has published a numerical comparison of many of the approaches for a suite of test cases [66].

6.1 Projection Scheme

Brackbill and Barnes [12] proposed using a Hodge-type projection to the magnetic field. This approach leads to a Poisson equation that must be solved each time the projection takes place:

$$\nabla^2 \phi = \nabla \cdot \mathbf{B} \tag{23a}$$

$$\mathbf{B}_{projected} = \mathbf{B} - \nabla \phi \tag{23b}$$

The resulting projected magnetic field is divergence-free on a particular numerical stencil, to the level of error of the solution of the Poisson equation. While it is not immediately obvious that the use of the projection scheme in conjunction with the fully conservative form of the MHD equations gives the correct weak solutions, Tóth has proven this to be the case [66]. The projection scheme has several advantages, including the ability to use standard software libraries for the Poisson solution, its relatively straightforward extension to general unstructured grids, and its robustness. It does, however, require solution of an elliptic equation at each projection step; this can be expensive, particularly on distributed-memory machines.

6.2 8-Wave Scheme

Powell [45,46] first proposed an approach based on the symmetrizable form of the MHD equations (eq. 8). In this approach, the source term on the right-hand side of eq. (8) is computed at each time step, and included in the update scheme. Discretizing this form of the equations leads to enhanced stability and accuracy, however, there is no stencil on which the divergence is identically zero. In most regions of the flow, the divergence source term is small. However, near discontinuities, it is not guaranteed to be small. In essence, the inclusion of the source term changes what would be a zero eigenvalue of the system to one whose value is u_n, the component of velocity normal to the interface through which the flux is computed. The scheme is typically referred to as the eight-wave scheme; the eighth wave corresponds to propagation of jumps in the normal component of the magnetic field.

The eight-wave scheme can be thought of as a hyperbolic or advective approach to controlling $\nabla \cdot \mathbf{B}$; symmetrizable form of the equations, eq. (8), are consistent with the passive advection of $\nabla \cdot \mathbf{B}/\rho$. The eight-wave scheme is computationally inexpensive, easy to add to an existing code, and quite robust. However, if there are regions in the flow in which the $\nabla \cdot \mathbf{B}$ source term (eq. 9) is large, the conservation errors can create problems.

6.3 Constrained Transport

Several approaches have been developed that have combined a Riemann-solver-based scheme with constrained-transport approach. The constrained-transport

approach of Evans and Hawley [19] treated the MHD equations in the gasdynamics/electromagnetic-split form of eq. (3a) through (3d). The grid used was a staggered one, and the $\nabla \cdot \mathbf{B} = 0$ constraint was met identically, on a particular numerical stencil.

Dai and Woodward [13] and Balsara and Spicer [1] modified the constrained-transport approach by coupling a Riemann-solver-based scheme for the conservative form of the MHD equations (eq. 5), with a constrained-transport approach for the representation of the magnetic field. In their formulations, this required two representations of the magnetic field: a cell-centered one for the Godunov scheme, and a face-centered one to enforce the $\nabla \cdot \mathbf{B} = 0$ condition. Tóth [66] subsequently showed that these formulations could be recast in terms of a single cell-centered representation for the magnetic field, through a modification to the flux function used.

Advantages of the conservative constrained-transport schemes include the fact that they are strictly conservative and that they meet the $\nabla \cdot \mathbf{B} = 0$ constraint to machine accuracy, on a particular stencil. The primary disadvantage of the constrained transport technique is the difficulty in extending them to general grids. Tóth and Roe [67] and Balsara [3] made some progress on this front; they developed divergence-preserving prolongation and restriction operators, allowing the use of conservative constrained-transport schemes on h-refined meshes. De Sterck generalized the constrained transport method to unstructured grids [15].

However, it can also be shown that the conservative constrained-transport techniques lose their $\nabla \cdot \mathbf{B}$-preserving properties if different cells are advanced at different physical time rates [68]. This rules out the use of local time-stepping (see Sect. 9.1). Thus, while for unsteady calculations the cost of the conservative constrained-transport approach is comparable to the eight-wave scheme, for steady-state calculations (where one would typically use local time-stepping), the cost can be prohibitive.

6.4 Diffusive Control

Some of the most recent work on the $\nabla \cdot \mathbf{B} = 0$ constraint has been related to modifying the eight-wave approach by adding a source term proportional to the gradient of $\nabla \cdot \mathbf{B}$ so that the the divergence satisfies an advection-diffusion equation, rather than a pure advection equation. Similar schemes have been used for the Maxwell equations [41]. This technique, referred to as diffusive control of $\nabla \cdot \mathbf{B}$, has the same advantages and disadvantages as the eight-wave approach. It is not strictly conservative, but appears to keep the level of $\nabla \cdot \mathbf{B}$ lower than the eight-wave approach does. In other recent work by Dedner et al. [14], a generalized Lagrange-multiplier method has been proposed, incorporating the projection approach, the eight-wave approach, and the diffusive-control approach into a single framework. In addition, they also introduced a hyperbolic correction of $\nabla \cdot \mathbf{B}$ errors.

7 Block-Based AMR on Cartesian Grids

Adaptive mesh refinement techniques that automatically adapt the computational grid to the solution of the governing PDEs can be very effective in treating problems with disparate length scales. Methods of this type avoid underresolving the solution in regions deemed of interest (e.g., high-gradient regions) and, conversely, avoid overresolving the solution in other less interesting regions (low-gradient regions), thereby saving orders of magnitude in computing resources for many problems. For typical solar wind flows, length scales can range from tens of kilometers in the near Earth region to the Earth-Sun distance (1.5×10^{11} m), and timescales can range from a few seconds near the Sun to the expansion time of the solar wind from the Sun to the Earth ($\sim 10^5$ s). The use of AMR is extremely beneficial and almost a virtual necessity for solving problems with such disparate spatial and temporal scales.

7.1 Adaptive Blocks

Borrowing from previous work by Berger and coworkers [5,6,8,9,10] and Quirk [47,48], and keeping in mind the desire for high performance on massively parallel computer architectures, a relatively simple yet effective block-based AMR technique has been developed and is used in conjunction with the finite-volume scheme described above. The method has some similarities with the block-based approaches described by Quirk and Hanebutte [48] and Berger and Saltzman [10]. Here the governing equations are integrated to obtain volume-averaged solution quantities within rectangular Cartesian computational cells. The computational cells are embedded in regular structured blocks of equal sized cells. The blocks are geometrically self-similar with dimensions $\tilde{\ell}_x \times \tilde{\ell}_y \times \tilde{\ell}_z$ and consist of $N_x \times N_y \times N_z$ cells, where $\tilde{\ell}_x$, $\tilde{\ell}_y$, and $\tilde{\ell}_z$ are the nondimensional lengths of the sides of the rectangular blocks and N_x, N_y, and N_z are even, but not necessarily all equal, integers. Typically, blocks consisting of anywhere between $4 \times 4 \times 4 = 64$ and $12 \times 12 \times 12 = 1728$ cells are used (see Fig. 1). Solution data associated with each block are stored in standard indexed array data structures. It is therefore straightforward to obtain solution information from neighboring cells within a block.

Computational grids are composed of many self-similar blocks. Although each block within a grid has the same data storage requirements, blocks may be of different sizes in terms of the volume of physical space that they occupy. Starting with an initial mesh consisting of blocks of equal size (i.e., equal resolution), adaptation is accomplished by the dividing and coarsening of appropriate solution blocks. In regions requiring increased cell resolution, a "parent" block is refined by dividing itself into eight "children" or "offspring." Each of the eight octants of a parent block becomes a new block having the same number of cells as the parent and thereby doubling the cell resolution in the region of interest. Conversely, in regions that are deemed overresolved, the refinement process is reversed, and eight children are coarsened and coalesced into a single parent block.

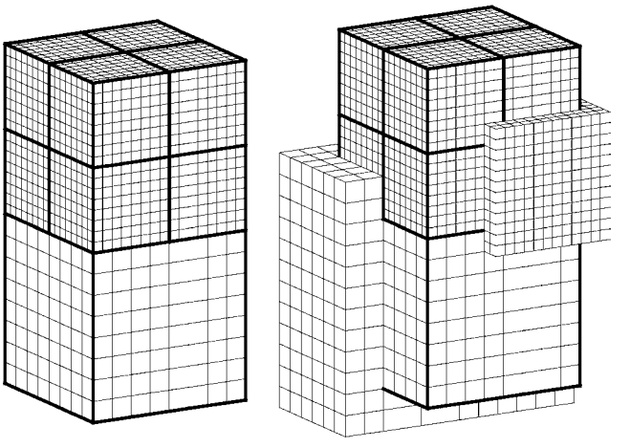


Fig. 1. (left) Self-similar blocks used in parallel block-based AMR scheme. (right) Self-similar blocks illustrating the double layer of ghost cells for both coarse and fine blocks.

In this way, the cell resolution is reduced by a factor of 2. Standard multigrid-type restriction and prolongation operators are used to evaluate the solution on all blocks created by the coarsening and division processes, respectively.

Two neighboring blocks, one of which has been refined and one of which has not, are shown in Fig. 1. Any of the blocks shown in Fig. 1 can in turn be refined, and so on, leading to successively finer blocks. In the present method, mesh refinement is constrained such that the cell resolution changes by only a factor of 2 between adjacent blocks and such that the minimum resolution is not less than that of the initial mesh.

In order that the update scheme for a given iteration or time step can be applied directly to all blocks in an independent manner, some additional solution information is shared between adjacent blocks having common interfaces. This information is stored in an additional two layers of overlapping "ghost" cells associated with each block as shown in Fig. 1. At interfaces between blocks of equal resolution, these ghost cells are simply assigned the solution values associated with the appropriate interior cells of the adjacent blocks. At resolution changes, restriction and prolongation operators, similar to those used in block coarsening and division, are employed to evaluate the ghost cell solution values. After each stage of the multistage time-stepping algorithm, ghost cell values are reevaluated to reflect the updated solution values of neighboring blocks. With the AMR approach, additional interblock communication is also required at interfaces with resolution changes to strictly enforce the flux conservation properties of the finite-volume scheme [5,6,8]. In particular, the interface fluxes computed on more refined blocks are used to correct the interface fluxes computed on coarser neighboring blocks so as to ensure that the fluxes are conserved across block interfaces.

7.2 Data Structure

A hierarchical tree-like data structure with multiple "roots," multiple "trees," and additional interconnects between the "leaves" of the trees is used to keep track of mesh refinement and the connectivity between solution blocks. This interconnected "forest" data structure is depicted in Fig. 2. The blocks of the initial mesh are the roots of the forest, which are stored in an indexed array data structure. Associated with each root is a separate "octree" data structure that contains all of the blocks making up the leaves of the tree which were created from the original parent blocks during mesh refinement. Each grid block corresponds to a node of the tree. Traversal of the multitree structure by recursively visiting the parents and children of solution blocks can be used to determine block connectivity. However, in order to reduce overhead associated with accessing solution information from adjacent blocks, the neighbors of each block are computed and stored directly, providing interconnects between blocks in the hierarchical data structure that are neighbors in physical space.

One of the advantages of the preceding hierarchical data structure is that it is relatively easy to carry out local mesh refinement at anytime during a calculation. If, at some point in a computation, a particular region of the flow is

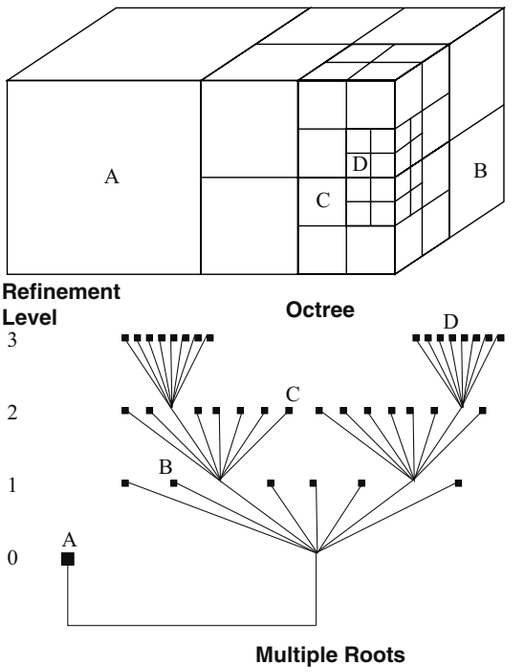


Fig. 2. Solution blocks of a computational mesh with three refinement levels originating from two initial blocks and the associated hierarchical multiroot octree data structure. Interconnects to neighbors are not shown.

deemed to be sufficiently interesting, better resolution of that region can be attained by refining the solution blocks in that region, without affecting the grid structure in other regions of the flow. Reducing the grid resolution in a region is equally easy. There is no need for completely remeshing the entire grid and recalculating block connectivity every time a mesh refinement is performed. Although other approaches are possible, in BATSRUS the coarsening and division of blocks are directed using multiple physics-based refinement criteria [43,44,46]. In particular, decisions as to when to refine or coarsen blocks are made based on comparisons of the maximum values of various local flow quantities determined in each block to specified refinement threshold values. Note that the refinement thresholds are dynamically adjusted so as to exercise some control over the total numbers of blocks, and hence cells, used in a calculation. We also note that other refinement criteria can also be used, such as a combination of estimated numerical errors.

An example illustrating the adaptation of the block-based Cartesian mesh to an evolving solution is shown in Fig. 3, which shows the grid at four different instances in time for an unsteady calculation and depicts both the solution blocks (thick lines) and computational cells (thin lines) of the evolving grid. As noted

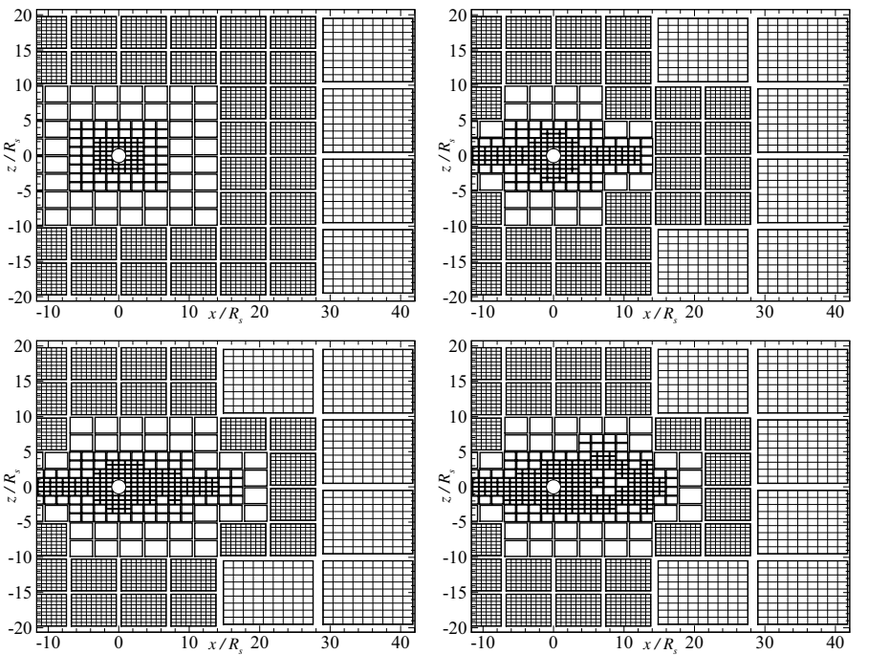


Fig. 3. Evolution of a computational mesh illustrating grid adaptation in response to changes in the numerical solution. Cross sectional cuts through a 3-D grid are shown for a solar wind calculation at four different instances in time. The computational cells are not shown for the smaller blocks.

above, each level of refinement in the grid introduces cells that are smaller by a factor of 2 in each dimension from those one level higher in the grid. Typically, calculations may have 10-15 levels of refinement; some calculations may have more than 20 levels of refinement. In the case of 20 levels of refinement, the finest cells on the mesh are more than one million times (2^{20}) smaller in each dimension than the coarsest cells. The block-based AMR approach described above has many similarities to the cell-based method proposed by De Zeeuw and Powell [16]. Although the block-based approach is somewhat less flexible and incurs some inefficiencies in solution resolution as compared to a cell-based approach, the block-based method offers many advantages over a cell-based technique when parallel implementations of the algorithms are considered and performance issues are taken into account. As will be discussed below, the block adaptation readily enables domain decomposition and effective load balancing and leads to low communication overhead between solution cells within the same block.

8 Parallel Implementation

The parallel block-based AMR solver was designed from the ground up with a view to achieving very high performance on massively parallel architectures. The underlying upwind finite-volume solution algorithm, with explicit time stepping, has a very compact stencil and is therefore highly local in nature. The hierarchical data structure and self-similar blocks make domain decomposition of the problem almost trivial and readily enable good load-balancing, a crucial element for truly scalable computing. A natural load balancing is accomplished by simply distributing the blocks equally amongst the processors. Additional optimization is achieved by ordering the blocks using the Peano-Hilbert space filling curve to minimize inter-processor communication. The self-similar nature of the solution blocks also means that serial performance enhancements apply to all blocks and that fine grain parallelization of the algorithm is possible. The parallel implementation of the algorithm has been carried out to such an extent, that even the grid adaptation is performed in parallel.

Other features of the parallel implementation include the use of FORTRAN 90 as the programming language and the message passing interface (MPI) library for performing the interprocessor communication. Use of these standards greatly enhances the portability of the code and leads to very good serial and parallel performance. The message passing is performed in an asynchronous fashion with gathered wait states and message consolidation.

Implementation of the algorithm has been carried out on Cray T3E supercomputers, SGI and Sun workstations, on Beowulf type PC clusters, on SGI shared-memory machines, on a Cray T3D, and on several IBM SP2s. BATSRUS nearly perfectly scales to 1,500 processors and a sustained speed of 342 GFlops has been attained on a Cray T3E-1200 using 1,490 PEs. For each target architecture, simple single-processor measurements are used to set the size of the adaptive blocks. The scaling of BATSRUS on various architectures is shown in Fig. 4.

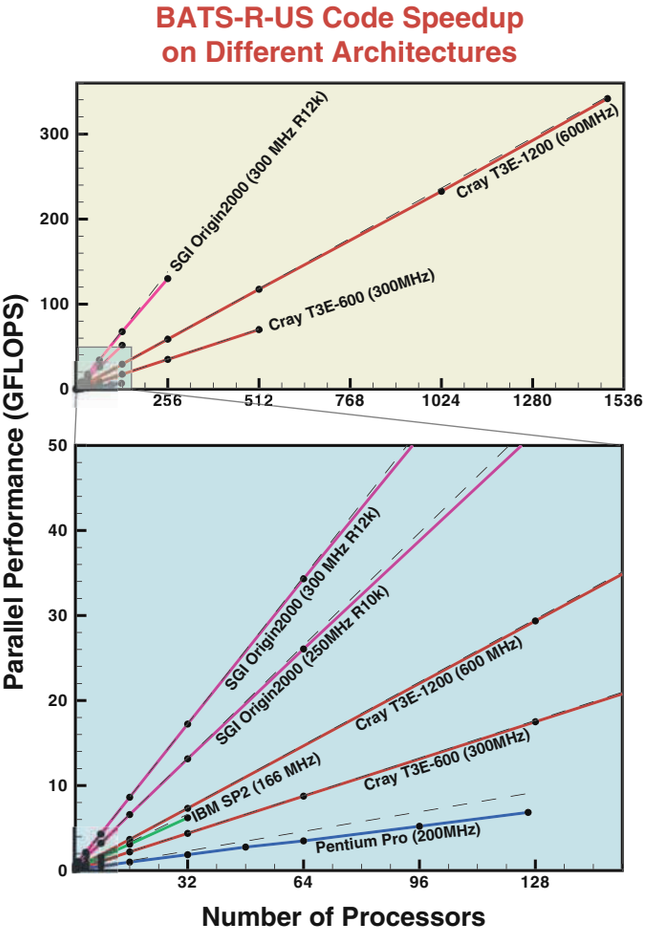


Fig. 4. Parallel speedup of BATSRUS on various architectures. Black dashed lines represent perfect scaling from single node performance.

9 Time to Solution

Since a major goal of global space plasma simulations is the creation of a predictive space weather tool, wallclock time to solution is a paramount issue. In particular, a predictive model must run substantially faster than real time. From the starting point – the observation of a solar event, to the ending point – post-processing the data from a simulation based on the initial conditions derived from the observations, a simulation must be accomplished rapidly to be of use.

The main limitation of the present generation of global space plasma codes is the explicit time stepping algorithm. Explicit time steps are limited by the Courant-Friedrichs-Lewy (CFL) condition, which essentially ensures that no information travels more than a cell size during a time step. This condition rep-

resents a non-linear penalty for highly resolved calculations, since finer grid resolution not only results in more computational cells, but also in smaller time steps.

In global MHD simulations of space plasmas the CFL condition is controlled by two factors: (1) the smallest cell size in the simulation, and (2) the fast magnetosonic speed in high magnetic field, low plasma density regions. In a typical magnetosphere simulation with a smallest cell size of about 0.25 R_{E} the CFL condition limits the time step to about 10^{-2} s. This small step is primarily controlled by the high fast magnetosonic speed (due to the high Alfvén speed) in the near-Earth region.

9.1 Local Time Stepping

In the local time stepping approach the time step in each cell of the computational domain is determined by the local stability condition. The flow variables in cell i are advanced from time step n to time step $n+1$ as

$$U_i^{n+1} = U_i^n + \Delta t_i^n \left(-\nabla \cdot \mathbf{F} + Q\right)_i \tag{24}$$

where the local time step is determined from the stability condition. Here U represents the conservative state vector, $\mathbf{F}$ is the flux diad and Q is the source term. In case of ideal MHD, the time step is determined by the CFL condition

$$\Delta t_i^n = C \frac{\Delta x_i}{c_i^{fast} + |u_i|} \tag{25}$$

where $C < 1$ is the Courant number and c_i^{fast} is the fast speed in cell i. In more than 1D the sum of the speeds in all directions should be taken in the denominator.

Note that this technique is different from "sub-cycling" when cells are advanced at the same physical time rate, but the number of time-steps taken by individual cells varies. For example, in adaptive grids it is customary to set the time-step to be inversely proportional to the size of the cell, so that a finer cell typically makes two half time-steps while the coarser cell makes only one full time-step. In this method the time-steps are still determined by a global stability condition as opposed to local time-stepping where time-steps are set on a cell-by-cell basis.

It is easy to see from (24) that the steady state solution, if it exists, satisfies

$$0 = \left(-\nabla \cdot \mathbf{F} + Q\right)_i \tag{26}$$

since in steady state $U_i^{n+1} = U_i^n$ and we can simplify with the time step Δt_i^n which is always a positive number. Consequently the steady state solution is independent of the time step, and it does not matter if it is local or global.

The above proof assumes that the steady state is fully determined by the boundary conditions. This is a non-trivial assumption, because the MHD equations are non-linear, and there is no mathematical theorem that would guarantee

the existance and uniquness of the steady state solution. In practice, magnetosphere simulations seem to converge to the same solution independent of the initial conditions or the time integration scheme.

The applicability of the local time-stepping technique in a given scheme primarily depends on the evolution of $\nabla \cdot \mathbf{B}$. In some methods even if $\nabla \cdot \mathbf{B} = 0$ initially, the numerical transients towards steady state will destroy this property if the local time stepping is applied. For instance, it can be shown that the constrained transport scheme cannot be combined with local time stepping.

9.2 Implicit Time-Stepping

In BATSRUS we have a number of time stepping algorithms implemented. The simplest and least expensive scheme is a multistage explicit time stepping, for which the time step is limited by the CFL stability condition. We have also implemented an unconditionally stable fully implicit time stepping scheme [65,36]. The second order implicit time discretization (BDF2) requires the solution of a non-linear system of equations for all the flow variables. This can be achived by the Newton-Krylov-Schwarz approach: a Newton iteration is applied to the nonlinear equations; a parallel Krylov type iterative scheme is used to solve the linear systems; the convergence of the Krylov solver is accelerated with a Schwarz type preconditioning. In our implementation the Krylov solver is BiCGSTAB, and a modified block incomplete LU (MBILU) preconditioner is applied on a block by block basis. Since every block has a simple Cartesian geometry, the preconditioner can be implemented very efficiently. The resulting implicit scheme requires about 20-30 times more CPU time per time step than the explicit method, but the physical time step can be 1,000 to 10,000 times larger. This implicit algorithm has a very good parallel scaling due to the Krylov scheme and the block by block application of the preconditioner.

In BATSRUS, we can combine explicit and implicit time stepping. Magnetosphere simulations include large volumes where the Alfvén speed is quite low (tens of km/s) and the local CFL number allows large explicit time steps (tens of seconds to several minutes). In these regions implicit time stepping is a waste of computational resources. Since the parallel implicit technique we use is fundamentally block based we only treat those blocks implicitly where the CFL condition would limit the explicit time step to less than the selected time step (typically ~ 10 s). Needless to say, this combined explicit-implicit time stepping represents more computational challenges (such as separate load balancing of explicit and implicit blocks). Overall, this solution seems to be a very promising option, but other potential avenues need to explored before one makes a final decision about the most efficient time-stepping algorithm for space MHD simulations. These questions will be discussed in an upcoming paper [68].

10 Ionosphere–Magnetosphere Coupling

The state of the magnetosphere is controlled by conditions in the solar wind and in the ionosphere. Solar wind conditions are imposed as boundary conditions at

the outer boundaries of the simulation domain. Even though the numerical procedures used to impose these boundary conditions can be quite sophisticated, the distant solar wind (in all directions) is assumed to be unaffected by the presence of the magnetosphere. This assumption makes the outer boundary conditions relatively simple (at least in principle).

The ionosphere–magnetosphere (I-M) coupling, on the other hand, is a highly non-linear two-way interaction which strongly affects the large-scale behavior of both domains. Self-consistent global magnetosphere models include some kind of dynamic ionosphere model which interacts with the magnetosphere and provides ionospheric boundary conditions actively responding to changing magnetospheric conditions.

While mass exchange between the ionosphere and the magnetosphere is undoubtedly of major importance, the dominant component of I-M coupling is a system of field-aligned currents (FACs) connecting the magnetosphere and the high-latitude ionosphere. These FACs carry momentum (electromagnetic stress) and energy (Poynting flux) along stretched magnetic field lines connecting the ionosphere and the magnetosphere. Self-consistent global magnetosphere models need to describe the generation and closure of these FACs through appropriate boundary conditions and embedded non-MHD models.

The most important current systems coupling the ionosphere and the magnetosphere are the so called Region 1 and Region 2 currents. Region 1 FACs, flowing near the open-closed magnetic field boundary, connect the magnetopause current to the ionosphere where they close through ionospheric Pedersen currents. Region 2 FACs flow along closed magnetic field lines and they connect to the ionosphere at lower magnetic latitudes than the Region 1 current. Region 2 currents are generated in the inner part of the plasma sheet and in the ring current region.

The generation of Region 2 currents can not be adequately described by an MHD model alone. The main physical process leading to the generation of these currents is the gradient and curvature drifts of hot magnetospheric particles. Since single-fluid MHD represents all particles with a single bulk velocity (obtained as a mass weighted average), it provides an inaccurate representation of the important drift processes (in reality electrons and ions drift in opposite directions, and ions with different energy and equatorial pitch-angle have very different drift velocities). As a result of this limitation of the single-fluid treatment, one needs to embed an accurate drift model into the global MHD code in order to get physically consistent Region 2 currents.

BATSRUS has been coupled to two ionosphere models to provide ionospheric current closure (the user can choose from the two available models) and a drift model to account for the generation and closure of FACs.

10.1 Electrostatic Ionosphere

The simplest I-M coupling procedure involves a height integrated electrostatic ionosphere model. In this case the MHD code has an inner boundary at a radius of $R_{\rm B}$ (for explicit time-stepping we typically use $R_{\rm B} = 2.5\,R_{\rm E}$). At this inner

boundary, the plasma density, temperature, and velocity are specified. In addition, the magnetic field is allowed to float (zero gradient of the normal component), so currents can flow along the boundary. The velocities which are imposed on the boundary are calculated in the ionosphere in a three step process:

1. field aligned currents are calculated from the curl of the magnetic field at $R_{\rm B} + 1\,R_{\rm E}$, and these are mapped down to the ionosphere,
2. a height-integrated ionospheric conductance pattern is generated and the ionospheric potential is calculated from the equation:

$$j_{\rm R}\left(R_{\rm E}\right) = \left[\nabla_{\perp} \cdot \left(\Sigma \cdot \nabla\psi\right)_{\perp}\right]_{R=R_{\rm E}} \tag{27}$$

   which describes the relationship between the height integrated conductance tensor, Σ, the ionospheric potential, ψ, and the radial component of the current, $j_{\rm R}$; and finally
3. the electric potential is mapped out along field lines to the inner boundary at $R_{\rm B}$ where electric fields and velocities are generated. The corotation velocity field is added to the ionosphere generated velocity field.

The details of our conductance model as well as some simulations results using this method are given in [52,53,55].

10.2 BATSRUS–TIEGCM Coupling [54]

As our physical understanding of the magnetosphere – ionosphere system increases, our need to have more accurate models of the coupled system also grows. Recently we developed a coupled model of the magnetosphere – ionosphere – thermosphere system. The model is based on BATSRUS [46] and the thermosphere – ionosphere – electrodynamics general circulation model (TIEGCM) [50] which has been used in numerous studies of the ionosphere - thermosphere system.

The TIEGCM solves for the thermospheric and ionospheric composition, temperature, and winds. It solves for mass mixing ratios of the neutral major species O_2, N_2, and O using full transport and chemistry, while the minor species $N(^2D)$, $N(^4S)$, NO, He, and Ar are obtained by assuming that they are in local equalibrium with the major species. For the ions, the O_2^+ dynamics are considered, while the species N_2^+, NO^+, and N are considered to be in local equalibrium with O_2^+. The TIEGCM is a full 3-dimensional code with 5° latitude by 5° longitude by 0.5 scale height altitude cells. There are 29 pressure levels in the model such that the simulation spans from $\sim$ 95 km to 650 km in altitude.

The electrodynamics within the TIEGCM focuses on the middle and lower latitudes, with a self consistent calculation of the interaction between the neutral winds, currents, and electric fields. At the high latitudes, an external electric field is assumed, and the neutral winds do not affect this field (i.e. there is no ion drag). At all latitudes, the neutral winds are coupled with ion flow through ion drag terms. A transition region exists where the model linearly scales the electric field between the high and middle latitude regions. The electrodynamics of the

TIEGCM are solved on a magnetic apex grid based on the 1985 International Geomagnetic Reference Field (IGRF) [51].

The main auroral electron precipitation is specified at high latitudes using an external model [57]. In addition to the main oval, a polar cap precipitation is specified as well as a spatially limited cusp precipitation. These precipitation patterns are used to generate 3-d ionization rates, and therefore strongly control the electron density at high latitudes.

The ionosphere is coupled to the MHD code in a similar manner as in the case of an electrostatic ionosphere (see section 10.1). The field aligned currents at 3.5 $R_{\rm E}$ are derived from $\nabla \times \mathbf{B}$. These currents are mapped down to the ionosphere using the intrinsic magnetic field within the code, and the currents are scaled by the ratio of the magnetic field in the ionosphere to that in the magnetosphere. The currents are combined with a conductance pattern to produce a potential pattern, which is mapped up to the magnetospheric inner boundary at 2.5 $R_{\rm E}$. Corotation velocities are determined and added to the convection velocities derived from the potential pattern. These velocities are applied as the inner boundary condition on the MHD solution.

BATSRUS provides the ionospheric potential pattern and electron precipitation pattern for TIEGCM. The electon precipitation is used to determine the nightside ionization in calculating the electron density. The electron density, combined with the densities of different neutral consituents, is subsequently used to determine the densities of different ion species. Collision rates between ions and neutrals are then calculated. These collision rates specify the acceleration of the neutral wind and the different conductivities.

In the low- and mid-latitude region, the neutral winds drive the dynamo calculation. This is done by taking a field-line integral of the neutral wind velocity multiplied by the conductivity. The divergence of this quantity is considered to be a quasi-neutral wind field aligned current, and is used for the calculation of the low- and mid-latitude dynamo in the TIEGCM. This quantity is also fed back into the MHD potential solver, where it is subtracted from the magnetospheric FACs. The neutral wind driven field aligned currents above 60° are considered in the calculation of the MHD potential, while those below are ignored.

The conductances are not completely consistent between the two models. The reasoning for not using the TIEGCM calculated conductances in the MHD model is that the grid size in the TIEGCM (5° latitude by 5° longitude) is too large. The auroral zone is ill represented by the TIEGCM. Because the particle precipitation is calculated by the MHD model, and the conductances derived from the Robinson formula [56] closely match those derived by models such as the TIEGCM, it was decided that it is better to use the high resolution conductances from the MHD model than the low resolution conductances from the TIEGCM. Once the resolution of the TIEGCM is increased, a more self-consitent coupling can occur.

The BATSRUS–TIEGCM coupling has been fully implemented and tested the first results obtained with the coupled model have also been reported [54].

10.3 BATSRUS–RCM Coupling [17]

The Rice Convection Model (RCM) calculates the dynamic behavior of the particles and the electric fields and currents in the Earth's inner magnetosphere (the region of closed magnetic lines). The physics of this region is complicated, because it contains overlapping particle distributions with a wide range of energies and characteristics. These different coexisting particle populations cannot be treated as a single fluid, because they all move differently. The RCM represents the particles in terms of 30-200 separate fluids. Its equations and numerical methods have been specifically designed for accurate treatment of the inner magnetosphere [31,29,71,18], including the flow of electric currents along magnetic field lines to and from the conducting ionosphere. The RCM does its primary calculations on a 2D grid on a spherical shell in the ionosphere. Values in the magnetosphere are computed by mapping out along magnetic field lines. The ionospheric grid is fine, typically about 0.5 degrees latitude in the auroral zone. The RCM computes these currents and the associated electric fields self-consistently.

The essential limitation of the RCM is that it only describes the inner magnetosphere: one needs to specify its connections to the outer magnetosphere and to the ionosphere and thermosphere through externally specified magnetic field models and particle source/sink boundary conditions. This can be done by coupling RCM to a global MHD code which dynamically provides these inputs and uses the results of the drift physics represented by RCM to improve the MHD solution.

RCM calculates the Region 2 field-aligned currents using the Vasyliunas formula [70]. This formula (which is also valid in the MHD limit) relates the field aligned current in slow flow regions to the cross product of the pressure gradient and the gradient of the magnetic flux tube volume. The pressure is calculated as a sum of partial pressures of the large number of "species," while the flux tube volume is calculated by integrating the flux tube cross secton ($\propto 1/B$) along magnetic field lines. In dynamic calculations (when the magnetic field is changing with time) the flux tube volume calculation is the most time consuming component of the entire simulation.

As a first step in coupling BATSRUS and RCM the Rice Convection Model was reformulated and a modern advection module was incorporated. Specifically the following changes were made:

1. Make the RCM code and its data structure compatible with BATSRUS.
2. Introduce MPI parallelization into RCM.
3. Introduce a new numerical scheme for solving the advection equations to achieve second-order accuracy and to handle time-dependent moving boundary conditions.
4. Improve other numerical algorithms (e.g., Birkeland current calculation.)

These changes in RCM already resulted in new physics, since the updated RCM code has more powerful capabilities than the original [60].

In the second step we introduced full, two-way coupling between BATSRUS and RCM. BATSRUS provides to RCM the full near-earth magnetic field (all closed field lines) as well as the flux-tube volume for the magnetic field line originating at each RCM grid point in the ionosphere. In addition, BATSRUS also provides the high-latitude ionospheric electric potential distribution and the particle sources at the outer boundaries at RCM. In return, RCM provides the hot plasma pressure averywhere in the closed field line region, which is used to to "nudge" the MHD pressure towards the RCM pressure (this is done through a source term).

The most computationally challenging aspect of this two-way coupling procedure is mapping between the 2D RCM and 3D MHD grids and the related calculation of the flux-tube volume for the RCM grid. This calculation is extremely time consuming if one uses standard field-line tracing routines and methods (in case of a reasonably refined grid this step might take hours of CPU time). We developed a block-based parallel ray-tracing algorithm which is about 10^4 times faster than the "brute force" method using a modest number of processors (about 32) [17]. This new algorithm makes the frequent two-way information exchange possible between BATSRUS and RCM.

The self-consistently coupled BATSRUS–RCM code has been recently tested and the results are being published [17]. The coupled code is capable describing the I-M coupling in a self-consistent manner, including the Region 1 and Region 2 FAC systems.

11 Applications

BATSRUS has been extensively applied to global numerical simulations of the inner heliosphere including CME propagation [24,26], the coupled terrestrial magnetosphere-ionosphere [23,61,62], and the interaction of the heliosphere with the interstellar medium [38]. In addition, it has also been successfully applied to a host of planetary problems ranging from comets [22,27], to Mercury [34], Venus [4], Mars [40], Saturn [28], to planetary satellites [33,35].

In this section we briefly summarize our most ambitious space weather simulation so far, in which we used BATSRUS to simulate an entire space weather event, from its generation at the Sun through the formation and evolution of a CME, to its interaction with the magnetosphere-ionosphere system [24,26]. In this simulation we resolved multiple spatial and temporal scales and took advantage of frequent grid refinements and coarsening to follow the CME through interplanetary space. The total number of cells varied between 800,000 and 2 million as the solution evolved. The simulation used 13 levels of grid refinement. The simulation ran faster than real time on a 512 node Cray T3E-600 supercomputer. This simulation demonstrates that we have the necessary experience to undertake the research outlined in this proposal.

Here we only show a few highlights of this simulation. The detailed results have been published in JGR-Space Physics [26].

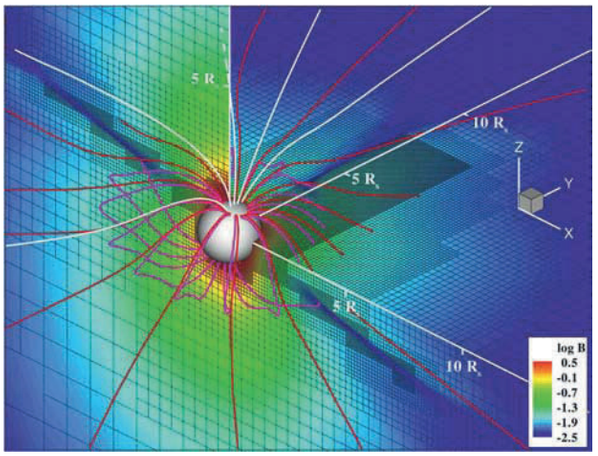


Fig. 5. 3D representation of the steady-state solar wind solution. The shading represents $\log |\mathbf{B}|$ in the (x, z)- and (x, y)-planes. The thin black lines are the computational mesh and the thick solid lines are magnetic field lines: grey denotes the last closed field lines, black is open field lines expanding to the interplanetary medium just above the heliospheric current sheet, and finally, white lines show open magnetic field lines in the (y, z)-plane.

A steady state solar wind was obtained in the corotating frame for a tilted rotating Sun. The intrinsic magnetic field was approximated by the superposition of a tilted (with respect to the rotation axis) octupole and dipole. Figure 5 depicts a three-dimensional representation of the predicted pre-event steady-state solar wind solution in the vicinity of the Sun. The narrow dark region shown in Fig. 5, which also coincides with regions of higher mesh refinement, corresponds to the beginning of the heliospheric current sheet. Due to the combined effects of magnetic tilt and solar rotation, the current sheet is tilted with respect to the rotation axis, and deformed, and resembles a "ballerina skirt."

Figure 6 shows a 3D representation of the magnetic field configuration 9 hours after the initiation of the CME. The density enhancement first leads to the "filling" of the closed magnetic field lines with additional plasma and subsequent expansion of the closed field line region. One can see that the closed field lines become greatly stretched by the outward moving plasma. This is due to the fact that the plasma β (the ratio of the kinetic and magnetic pressures) is quite large and the magnetic field is "carried" by the outward moving plasma. We also note the decrease of magnetic field strength behind the leading edge of the outward moving disturbance.

The dynamic response of the global magnetosphere to the changing solar wind conditions produced by the density-driven CME was also computed as part of this simulation. The global magnetospheric configuration for quiet-time southward IMF conditions is shown in Fig. 7. During the event the solar wind velocity remained nearly radial with the speed gradually decreasing from about

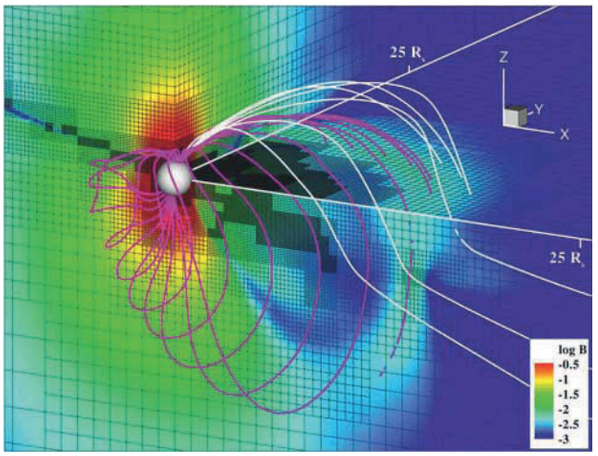


Fig. 6. 3D representation of magnetic field lines 9 hours after the initiation of a CME. Grayscale represents $\log(B)$, white lines are open magnetic field lines, grey lines represent magnetic field lines with both ends connected to the Sun.

550 km/s to about 450 km/s. The solar wind dynamic pressure increased from its pre-CME value of 2.25 nP (at $t = 72$ hrs) to 4.6 nP at the peak of the event.

The ionospheric potential and convection patterns also change during the CME event. The ionospheric convection shows the two-cell pattern of ionospheric convection typical for southward-type IMF conditions. The convection pattern is also "twisted" due to the presence of a non-zero IMF B_y component. The most important change in the ionosphere is the doubling of the cross-cap potential drop from 30 kV at 70.5h to 60 kV some 27 hours later.

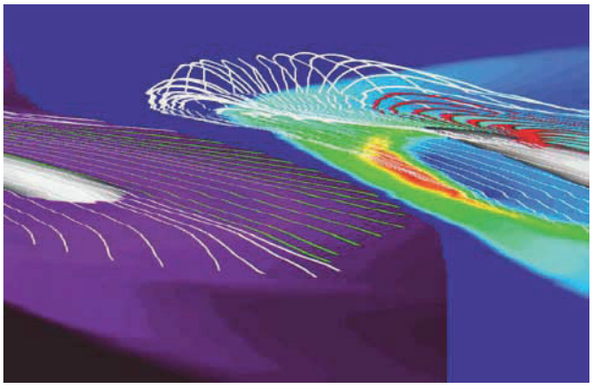


Fig. 7. 3D representation of the last closed terrestrial field lines for southward IMF conditions. White field lines form the dayside magnetopause, while black ones map to the magnetotail. The greyscale represents normalized thermal pressure.

Overall, this simulated space weather event was not very geoeffective. It is expected that we will be able to generate more geoeffective CMEs with the help of more realistic explosive event generation modules. This simulation, however, demonstrates the present capabilities of BATSRUS.

Acknowledgements

This work was supported by DoD MURI grant F49620-01-1-0359, NSF KDI grant NSF ATM-9980078, NSF CISE grant ACI-9876943, and NASA AISRP grant NAG5-9406. G. Tóth is partially supported by the Education Ministry of Hungary (grant No. FKFP-0242-2000). We also acknowledge the contributions of the developers of TIEGCM and RCM to the coupled magnetosphere model. In particular, the contributions of Ray Roble, Stanislav Sazykin and Richard Wolf are acknowledged and appreciated.

References

1. D. S. Balsara and D. S. Spicer: J. Comput. Phys., **149**, 270 (1999)
2. D. S. Balsara and D. S. Spicer: J. Comput. Phys., **148**, 133 (1999)
3. D. Balsara: J. Comput. Phys., **174**, 614 (2001)
4. R. Bauske, A. F. Nagy, T. I. Gombosi, D. L. De Zeeuw, K. G. Powell and J. G. Luhmann: J. Geophys. Res., **103**, 23,625 (1998)
5. M. J. Berger: 'Adaptive Mesh Refinement for Hyperbolic Partial Differential Equations'. Interanl report at Stanford Univ., Stanford, Calif., 1982
6. M. J. Berger: J. Comput. Phys., **53**, 484 (1984)
7. M. J. Berger and A. Jameson: AIAA Journal, **32**, 561 (1985)
8. M. J. Berger and P. Colella: J. Comput. Phys., **82**, 67 (1989)
9. M. J. Berger and R. J. LeVeque: 'An adaptive Cartesian mesh algorithm for the Euler equations in arbitrary geometries'. In *Proc. 9th AIAA Computational Fluid Dynamics Conference* (AIAA Paper No. 89-1930, Buffalo, NY 1989)
10. M. J. Berger and S. Saltzman: "Appl. Numer. Math.", **14**, 239 (1994)
11. Boris, J. P.: 'A physically motivated solution of the Alfvén problem'. Internal report at Naval Research Laboratory, NRL Memorandum Report 2167, Washington, D.C., 1970
12. J.U. Brackbill and D.C. Barnes: J. Comput. Phys., **35**, 426 (1980)
13. W. Dai and P. R. Woodward: J. Comput. Phys., **142**, 331 (1998)
14. Dedner, A., F. Kemm, D. Kröner, C.-D. Munz, T. Schnitzer and M. Wesenberg: J. Comput. Phys., **175**, 645 (2002)
15. H. De Sterck: 'Multi-Dimensional Upwind Constrained Transport on Unstructured Grids for 'Shallow Water' Magnetohydrodynamics'. In *Proceedings of AIAA 15th Computational Dynamics Conference*, (AIAA-2001-2623, 2001)
16. D. L. De Zeeuw and K. G. Powell: J. Comput. Phys., **104**, 55 (1993)
17. D. L. De Zeeuw, S. Sazykin, A. J. Ridley, G. Tóth, T. I. Gombosi, K. G. Powell and R. A. Wolf: J. Geophys. Res., (2002) "in preparation",
18. Erickson, G. M., R. W. Spiro and R. A. Wolf: J. Geophys. Res., **96**, 1633 (1991)
19. C. R. Evans and J. F. Hawley: Astrophys. Journ., **332**, 659 (1988)
20. Fedder, J. A., S. P. Slinker, J. G. Lyon and R. D. Elphinstone: J. Geophys. Res., **100**, 19,083 (1995)

21. S. K. Godunov: 'Symmetric form of the equations of magnetohydrodynamics' (in Russian). In: *Numerical Methods for Mechanics of Continuum Medium*, "Siberian Branch of USSR Acad. of Sci.", **1**, 26, Novosibirsk, 1972
22. T. I. Gombosi, D. L. De Zeeuw, R. M. Häberli and K. G. Powell: J. Geophys. Res., **101**, 15233 (1996)
23. T. I. Gombosi, D. L. De Zeeuw, C. P. T. Groth, K. G. Powell and P. Song: The length of the magnetotail for northward IMF: Results of 3D MHD simulations, In *Physics of Space Plasmas* ed. by T. Chang and J. R. Jasperse, (MIT Press, Cambridge, Mass., 1998) pp. 121-128
24. T. I. Gombosi, D. L. De Zeeuw, C. P. T. Groth, K. G. Powell and Q. F. Stout: J. Atmos. Solar-Terr. Phys., **62**, 1515 (2000)
25. T. I. Gombosi, G. Tóth, D. L. De Zeeuw, K. C. Hansen, K. Kabin and K. G. Powell: J. Comput. Phys., **177**, 176 (2002)
26. C. P. T. Groth, D. L. De Zeeuw, T. I. Gombosi and K. G. Powell: J. Geophys. Res., **105**, 25,053 (2000)
27. R. M. Häberli, T. I. Gombosi, D. L. DeZeuuw, M. R. Combi and K. G. Powell: Science, **276**, 939 (1997)
28. K. C. Hansen, T. I. Gombosi, D. L. DeZeeuw, C. P. T. Groth and K. G. Powell: Adv. Space Res., **26**, 1681 (2000)
29. Harel, M., R. A. Wolf, P. H. Reiff, R. W. Spiro, W. J. Burke, F. J. Rich and M. Smiddy: J. Geophys. Res., **86**, 2217 (1981)
30. A. Harten: J. Comput. Phys., **49**, 357 (1983)
31. R. K. Jaggi and R. A. Wolf: J. Geophys. Res., **78**, 2842 (1973)
32. P. Janhunen: J. Comput. Phys., **160**, 649 (2000)
33. K. Kabin, T. I. Gombosi, D. L. DeZeeuw, K. G. Powell and P. L. Israelevich: J. Geophys. Res., **104**, 2451 (1999)
34. K. Kabin, T. I. Gombosi, D. L. De Zeeuw and K. G. Powell: "Icarus", **143**, 397 (2000)
35. K. Kabin, M. R. Combi, T. I. Gombosi, D. L. De Zeeuw, K. C. Hansen and K. G. Powell: Planet. Space Sci., **49**, 337 (2001)
36. R. Keppens, G. Tóth, M. A. Botchev and A. van der Ploeg: "Int. J. for Num. Meth. in Fluids", **30**, 335 (1999)
37. P. D. Lax and B. Wendroff: Systems of Conservation Laws, "Communications on Pure and Applied Mathematics", **13**, 217–237 (1960)
38. T. J. Linde, T. I. Gombosi, P. L Roe, K. G. Powell and D. L. De Zeeuw: J. Geophys. Res., **103**, 1889 (1998)
39. T. J. Linde: A Three-Dimensional Adaptive Multifluid MHD Model of the Heliosphere", school = "Univ. of Mich.", Ann Arbor, month = may, 1998
40. Y. Liu, A. F. Nagy, C. P. T. Groth, D. L. De Zeeuw, T. I. Gombosi and K. G. Powell: Geophys. Res. Lett., **26**, 2689 (1999)
41. B. Marder: J. Comput. Phys., **68**, 48 (1987)
42. T. Ogino and R. J. Walker: Geophys. Res. Lett., **11**, 1018 (1984)
43. H. Paillère, K. G. Powell and D. L. De Zeeuw: 'A wave-model based refinement criterion for adaptive-grid computation of compressible flows'. In: *30th AIAA Aerospace Sciences Meeting* (AIAA-92-0322, Reno, Nevada, 1992)
44. K. G. Powell, P. L. Roe and J. Quirk: 'Adaptive-mesh algorithms for computational fluid dynamics'. In: *Algorithmic Trends in Computational Fluid Dynmaics* ed. by M. Y. Hussaini, A. Kumar and M. D. Salas (SPRINGER-VERLAG, New York, 1993) pp. 303-337

45. K. G. Powell: 'An approximate Riemann solver for magnetohydrodynamics (that works in more than one dimension)'. Internal report at Inst. for Comput. Appl. in Sci. and Eng., NASA Langley Space Flight Center, Hampton, Va., 94-24, 1994
46. K. G. Powell, P. L. Roe, T. J. Linde, T. I. Gombosi and D. L. De Zeeuw: J. Comput. Phys., **154**, 284 (1999)
47. J. J. Quirk: 'An Adaptive Grid Algorithm for Computational Shock Hydrodynamics'. Internal report at Cranfield Inst. of Technol., Cranfield, England", 1991
48. J. J. Quirk and U. R. Hanebutte: 'A parallel adaptive mesh refinement algorithm'. Interanl; report at ICASE, 93-63, 1993
49. J. Raeder, R. J. Walker and M. Ashour-Abdalla: Geophys. Res. Lett., **22**, 349 (1995)
50. Richmond, A. D., E. C. Ridley and R. G. Roble: Geophys. Res. Lett., **19**, 601 (1992)
51. Richmond, A. D.: J. Geomagn. Geoelectr., **47**, 191 (1995)
52. A. J. Ridley, D. L. De Zeeuw, T. I. Gombosi and K. G. Powell: J. Geophys. Res., **106**, 30,067 (2001)
53. A. J. Ridley, K. C. Hansen, G. Tóth, D. L. De Zueew, T. I. Gombosi and K. G. Powell: J. Geophys. Res., **107** (2002) "in press",
54. A. J. Ridley, T. I. Gombosi, D. L. De Zeeuw, C. R. Clauer and A. D. Richmond: J. Geophys. Res., **107** (2002) "in preparation"
55. A. J. Ridley, T. I. Gombosi and D. L. De Zeeuw: J. Geophys. Res. **107** (2002) "in preparation",
56. R.M. Robinson, R.R. Vondrak, K. Miller, T. Dabbs and D.A. Hardy: J. Geophys. Res., **92**, 2565 (1987)
57. R. G. Roble and E. C. Ridley: Ann. Geophys., **5A**, 369 (1987)
58. P. L. Roe: J. Comput. Phys., **43**, 357 (1981)
59. P. L. Roe and D. S. Balsara: SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS J. APPL. MATH., **56**, 57 (1996)
60. S. Sazykin and R. A. Wolf and R. W. Spiro and T. I. Gombosi and D. L. De Zeeuw and M. F. Thomsen: Geophys. Res. Lett., (2002) "in preparation"
61. P. Song, T.I. Gombosi, D.L. De Zeeuw, K.G. Powell and C. P. T. Groth: Planet. Space Sci., **48**, 29 (2000)
62. P. Song, D. L. De Zeeuw, T. I. Gombosi, C. P. T. Groth and K. G. Powell: J. Geophys. Res., **104**, 28,361 (1999)
63. J. M. Stone and M. L. Norman: Astrophys. J. Suppl., **80**, 791 (1992)
64. T. Tanaka: J. Comput. Phys., **111**, 381 (1994)
65. G. Tóth, R. Keppens and M. A. Botchev: Astron. Astrophys., **332**, 1159 (1998)
66. G. Tóth: J. Comput. Phys., **161**, 605 (2000)
67. G. Tóth and P. L. Roe: J. Comput. Phys., (2001) "submitted"
68. G. Tóth, D. L. De Zeeuw, T. I. Gombosi, K. C. Hansen, K. G. Powell, A. J. Ridley, P. L. Roe and I. V. Sokolov: J. Geophys. Res., **107** (2002) "in preparation"
69. B. van Leer: J. Comput. Phys., **32**, 101, (1979)
70. V. M. Vasyliunas: 'Mathematical models of magnetospheric convection and its coupling to the ionosphere'. In: *Particles and fields in the magnetosphere* ed. by B. M. McCormack (D. Reidel Publishing, Dordrecht, Holland 1970) pp. 60-71
71. R. A. Wolf: 'The quasi-static (slow-flow) region of the magnetosphere'. In: *Solar Terrestrial Physics*, ed. by R. L. Carovillano and J. M. Forbes, (D. Reidel Publishing, Hingham, MA 1983) pp. 303-368

Finite Volume TVD Schemes for Magnetohydrodynamics on Unstructered Grids

Takashi Tanaka

Communications Research Lab., Koganei-shi, Tokyo 184-8795, Japan

Abstract. A three-dimensional (3-D) high-resolution magnetohydrodynamic (MHD) simulation scheme is developed on unstructured grid systems to solve the complex-system problems in space science and space weather in which numerical difficulties arise from inhomogeneity due to strong background potential fields, inclusion of multi-species ions, and formations of shocks and discontinuities. The ideal MHD equations are extended to the 9-component MHD equations for multi-component ions and modified so as to avoid a direct inclusion of background potential field in dependent variables through the use of new variables. The numerical scheme adopts the finite volume method (FVM) with an upwinding numerical flux based on the linearized Riemann solver. Upwindings on unstructured grid systems are realized from the fact that the MHD equations are symmetric with respect to the rotation of the space. Despite the modifications of the equation system, the eigenvectors in the mode-synthesis matrix necessary for the evaluation of the upwinding numerical flux can still be written analytically. To get a higher order of accuracy, the upwinding flux is extended to the third-order total variation diminishing (TVD) numerical flux in the calculation of FVM, through the monotonic upstream scheme for conservation laws (MUSCL) approach and Van Leer's differentiable limiter. Three numerical examples are given in order to show the efficiency of the above scheme.

1 Introduction

Recently, the magnetohydrodynamic (MHD) simulations are widely applied to many problems in space science with a great success [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12]. A rapid development of supercomputers in computational speed and memory size gives a conviction for further developments in this fruitful area. In these studies, the developments of supercomputers and numerical schemes are like the two wheels of a cart. In order to apply the MHD simulations to the problems having more complex configurations, further improvements of numerical schemes are also unavoidable. In this paper, we develop a numerical MHD scheme that enables an exact treatment of multi-scale space plasma including multi-component ions and strong background potential field, with an excellent capturing of shocks and discontinuities.

In the space science, we must always study the complex systems which are controlled by the coupling processes between different regions having quite different characteristics. Auroral physics is a typical example of this kind of problem [3], [5]. The main process controlling this problem is the coupling effects that

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 275–295, 2003.

occur between the magnetosphere and the ionosphere. The characteristic length and time scales of these two regions are vastly different. If numerical MHD simulations are applied to these problems, then it becomes necessary to assign different grid point densities to each region, in order to facilitate the simultaneous treatment of the different regions. Therefore, it is unavoidable to adopt an unstructured grid system. The ordinary finite difference method (FDM) which is widely used to solve differential equations numerically is no longer applicable on an unstructured grid system and so here we consider the finite volume method (FVM) based on the flux conservation law [13].

High-speed flowing plasma that is frequently treated in space science tends to form shocks and discontinuities. The use of standard numerical schemes of second-order accuracy (e.g. the Lax-Wendroff method) generates spurious oscillations at high gradients. Therefore, there is a need to use more advanced schemes that can adequately represent these shocks and discontinuities. An outstanding approach is to evaluate numerical flux in the FVM from an upwinding method based on the linearized Riemann solver [14] [15] [16]. To get a higher order of accuracy, the upwinding scheme for numerical flux is extended to the total variation diminishing (TVD) scheme. Among many TVD schemes, a third-order TVD scheme based on the monotonic upstream scheme for conservation laws (MUSCL) approach is considered in this paper [17] [18]. With this numerical flux, excellent shock-capturing is enabled along with stable and highly-accurate computations. The eigenvalues and eigenvectors of the MHD flux Jacobian matrix necessary for the upwinding calculations are derived from the well-known Alfven, fast and slow velocities [14]. The calculation of eigenvectors is done with special care when wave propagations become parallel or perpendicular to the ambient magnetic field, because degenerations of eigenvalues occur in these cases [14] [19].

Another problem in space science is that many planets and stars treated in the complex-system simulation have a strong dipole magnetic field generated in their interior regions. In the case of the earth, the magnitude of the dipole magnetic field is about 30000 nT in the ionospheric region near the Earth, while it diminishes rapidly in the magnetosphere to about 10 nT. Therefore, the magnitude of the intrinsic magnetic field varies over a wide range in the whole treating region, in the problem of magnetosphere-ionosphere (M-I) couplings. On the other hand, the variable components of magnetic field, which are calculated from the MHD equations, exhibit a similar magnitude over the whole region. As a result, the ratio of variable to intrinsic components of the magnetic field becomes extremely small in the ionospheric region. These situations give a difficulty in the numerical study of the coupling process between two different regions. Especially, severe difficulties appear in the energy equation. However, this difficulty due to the wide range in the ratio of variable to internal magnetic fields can be avoided from the fact that intrinsic magnetic field includes only potential components. Thus, it becomes important to construct the MHD calculations suppressing the direct inclusion of the intrinsic component of the magnetic field, as dependent

variables [3] [19]. In this paper, therefore, a modified equation system is used to cope with such a problem, changing dependent variables.

In order to apply the MHD simulation to more complicated problems in space science, an additional improvements is imposed on the original MHD equations. Since space plasma does not always consist of single ion specie, sometimes we must treat the plasma that includes multi-component ions with the source and sink [9]. The ordinary 8-component MHD equations can be extended to the 9-component MHD equations for 2-component plasma.

It is shown in this paper that the equation system with the above modifications can still be written in the conservation form and can also be treated numerically through the FVM with the upwinding TVD flux. The eigenvalues and eigenvectors necessary to construct a TVD scheme are calculated for the 9-component MHD equations with modified variables. Construction of the scheme is seen in Sects. 2, 3, 4, and 5. In Sect. 6, a brief comment is given about the suitability of the present scheme for parallel computation. To show the feasibility of the scheme, three numerical examples are shown in Sects. 7, 8, and 9. The first example in Sect. 7 shows an excellent ability of present scheme for the capturing of shocks and discontinuities. The second example in Sect. 8 demonstrates the applicability of the present scheme to multi-component plasma. The last example in Sect. 9 adopts the present scheme to the space weather problem that includes a potential magnetic field, and shows the capability of low-noise calculation even in the low-β region.

2 TVD Schemes for Hyperbolic Equations

The history of the development of numerical schemes for hyperbolic conservation law is long and rich [20]. The idea that stable computation can be accomplished through an approximate dissipation term was used in early schemes. The advection equation, the most simple hyperbolic equation, can be written in a one dimensional coordinate system (x, t) as

$$\frac{\partial u}{\partial t} + C\frac{\partial u}{\partial x} = 0. \tag{1}$$

Where C is an advection speed. Let u_i^n be the numerical solution of (1) at $x = i \triangle x$ and $t = n \triangle t$. Then, the explicit time integration of this equation is

$$\frac{u_i^{n+1} - u_i^n}{\triangle t} = -C\frac{u_{i+1}^n - u_{i-1}^n}{2\triangle x} + \frac{d}{2}\frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{\triangle t}. \tag{2}$$

The stability condition for equation (2) is $1 > d > k^2$, with $k = |C|\triangle t/\triangle x$. Friedrichs-Lax, Godunov, and Lax-Wendroff schemes are obtained automatically by selecting $d = 1, d = k = |C\triangle t/\triangle x|$, and $d = k^2 = (C\triangle t/\triangle x)^2$, respectively. Out of these schemes only the Lax-Wendroff scheme has the second order accuracy.

Modern shock-capturing schemes add only enough dissipation in small localized regions to eliminate numerical oscillations. These schemes enforce some

constraint on the problem, usually that the solutions be TVD. The Godunov scheme can be modified to

$$u_i^{n+1} - u_i^n = C_{i+1/2}^-(u_{i+1}^n - u_i^n) - C_{i-1/2}^+(u_i^n - u_{i-1}^n), \tag{3}$$

with

$$C_{i+1/2}^- = (|C| - C)\triangle t/2\triangle x, C_{i-1/2}^+ = (|C| + C)\triangle t/2\triangle x. \tag{4}$$

These coefficients satisfy the condition

$$0 < C_{i+1/2}^- + C_{i-1//2}^+ < 1, 0 < C_{i+1/2}^-, C_{i-1//2}^+, \tag{5}$$

and the total variation at step $n+1$ under this condition is

$$\begin{aligned}
&TV(u^{n+1}) = \sum |u_{i+1} + C_{i+3/2}^-(u_{i+2} - u_{i+1}) - C_{i+1/2}^+(u_{i+1} - u_i) \\
&-u_{i+1} - C_{i+1/2}^-(u_{i+1} - u_i) + C_{i-1/2}^+(u_i - u_{i-1})| \\
&\leq \sum C_{i+3/2}^-|u_{i+2} - u_{i+1}| \\
&+ \sum (1 - C_{i+1/2}^+ - C_{i+1/2}^-)|u_{i+1} - u_i| \\
&+ \sum C_{i-1/2}^+|u_i - u_{i-1}| \\
&= \sum |u_{i+1} - u_i|.
\end{aligned} \tag{6}$$

Thus, the TVD is satisfied for the Godunov scheme. This sccheme is a first-order upwind TVD scheme.

In order to obtain a higher-order TVD scheme, we write equation (2) in the flux formula

$$\frac{u_i^{n+1} - u_i^n}{\triangle t} = -\frac{F_{i+1/2}^n - F_{i-1/2}^n}{\triangle x}, \tag{7}$$

then the Godunov and Lax-Wendroff fluxes in the case of $C > 0$ are

$$F_{i+1/2}^G = Cu_i, \tag{8}$$

and

$$F_{i+1/2}^L = Cu_i + C/2 \cdot (1-k)(u_{i+1} - u_i), \tag{9}$$

respectively. Equation (8) shows the fact that the Godunov scheme is an upwind scheme. By combining Godunov and Lax-Wendroff fluxes as

$$F_{i+1/2} = (1 - B_{i+1/2})F_{i+1/2}^G + B_{i+1/2}F_{i+1/2}^L, \tag{10}$$

one can obtain

$$\frac{u_i^{n+1} - u_i^n}{u_{i-1}^n - u_i^n} = k - \frac{1}{2}k(1-k)B_{i-1/2} + \frac{1}{2}k(1-k)\frac{B_{i+1/2}}{r_i}, \tag{11}$$

with

$$r_i = (u_i - u_{i-1})/(u_{i+1} - u_i). \tag{12}$$

In order for the scheme to be TVD, the left hand side of equation (11) must be between 0 and 1 and the resulting sufficient condition obtained after some simple algebra is the following,

$$-\frac{2}{k} < -2 < B_{i-1/2} - \frac{B_{i+1/2}}{r_i} < 2 < \frac{2}{1-k}, \tag{13}$$

which reduces to

$$0 < B_{i+1/2}, B_{i+1/2}/r_i < 2. \tag{14}$$

Commonly used limiters which give the constraint for the TVD scheme satisfy this condition.

For a coupled nonlinear equation system

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} = 0, \mathbf{A} = \frac{\partial \mathbf{f}}{\partial \mathbf{u}}, \tag{15}$$

the Riemann problem is solved to evaluate the upwind flux. From the flux Jacobian matrix $\mathbf{A}$, its eigenvalues λ_k, and eigenvectors $\mathbf{r}_k$, $\mathbf{u}_{i+1} - \mathbf{u}_i$ can be expanded as

$$\mathbf{u}_{i+1} - \mathbf{u}_i = \sum C_k \mathbf{r}_k, C_k = \mathbf{r}_k^{-1} \triangle \mathbf{u}, \mathbf{A}_{i+1/2} \mathbf{r}_k = \lambda_k \mathbf{r}_k, \tag{16}$$

then, the upwind flux can be evaluated as

$$\begin{aligned}
&\mathbf{F}_{i+1/2} = \mathbf{f}(\mathbf{u}^R_{i+1/2}) = \mathbf{f}(\mathbf{u}_i + \sum\nolimits^{-} C_k \mathbf{r}_k) = \mathbf{f}(\mathbf{u}_i) + \mathbf{A} \sum\nolimits^{-} C_k \mathbf{r}_k \\
&= \mathbf{f}(\mathbf{u}_i) + \sum\nolimits^{-} \lambda_k C_k \mathbf{r}_k = \mathbf{f}_i/2 + (\mathbf{f}_{i+1} - \sum \lambda_k C_k \mathbf{r}_k)/2 + \sum\nolimits^{-} \lambda_k C_k \mathbf{r}_k \\
&= (\mathbf{f}_i + \mathbf{f}_{i+1})/2 - (\sum\nolimits^{+} \lambda_k C_k \mathbf{r}_k - \sum\nolimits^{-} \lambda_k C_k \mathbf{r}_k)/2 \\
&= (\mathbf{f}_i + \mathbf{f}_{i+1})/2 - (\sum |\lambda_k| C_k \mathbf{r}_k)/2 \\
&= (\mathbf{f}_i + \mathbf{f}_{i+1})/2 - (\mathbf{R}|\mathbf{\Lambda}|\mathbf{R}^{-1} \triangle \mathbf{u})/2.
\end{aligned} \tag{17}$$

The accuracy of the scheme increases by considering the interpolationof dependent variables. An example is the MUSCL interpolation in which i and j are replaced by L and R as

$$\mathbf{u}_L = \mathbf{u}_i + \mathbf{s}_i\{(\mathbf{1} - \mathbf{s}_i/3)(\mathbf{u}_i - \mathbf{u}_{i-1}) + (\mathbf{1} + \mathbf{s}_i/3)(\mathbf{u}_{i+1} - \mathbf{u}_i)\}/4, \tag{18}$$

$$\mathbf{u}_R = \mathbf{u}_{i+1} - \mathbf{s}_{i+1}\{(\mathbf{1} + \mathbf{s}_{i+1}/3)(\mathbf{u}_{i+1} - \mathbf{u}_i) + (\mathbf{1} - \mathbf{s}_{i+1}/3)(\mathbf{u}_{i+2} - \mathbf{u}_{i+1})\}/4, \tag{19}$$

where $\mathbf{s}_i$ is a limiter at the grid point i.

Assuming first-order accuracy in implicit terms, we can also obtain an implicit MUSCL scheme as

$$\begin{aligned}&\frac{\triangle x}{\triangle t}\triangle \mathbf{u}_i^{n+1} + \frac{\partial \mathbf{F}_{i+1/2}}{\partial \mathbf{u}_{i+1}}\triangle \mathbf{u}_{i+1}^{n+1} + \frac{\partial \mathbf{F}_{i+1/2}}{\partial \mathbf{u}_i}\triangle \mathbf{u}_i^{n+1}\\ &-\frac{\partial \mathbf{F}_{i-1/2}}{\partial \mathbf{u}_i}\triangle \mathbf{u}_i^{n+1} - \frac{\partial \mathbf{F}_{i-1/2}}{\partial \mathbf{u}_{i-1}}\triangle \mathbf{u}_{i-1}^{n+1}\\ &= -(\mathbf{F}_{i+1/2}^n - \mathbf{F}_{i-1/2}^n),\end{aligned} \tag{20}$$

with a notation $\triangle \mathbf{u}_i^{n+1} = \mathbf{u}_i^{n+1} - \mathbf{u}_i^n$.

3 Finite Volume TVD Scheme for 9-Component MHD Equations

In space plasma simulation, sometimes we must treat multiple plasma components. The ideal MHD equations can be extended to the 9-component MHD equations that treat 2-component plasma. The nondimensional conservation-law form for these equations can be written in the Cartesian coordinate system (x,y,z,t) as

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{u})}{\partial x} + \frac{\partial \mathbf{G}(\mathbf{u})}{\partial y} + \frac{\partial \mathbf{H}(\mathbf{u})}{\partial z} = \mathbf{S}, \tag{21}$$

where the dependent variables are $\mathbf{u} = (\rho, \mathbf{m}, \mathbf{B}, U, \rho_2)^T$ and $\mathbf{F}, \mathbf{G}, \mathbf{H}$, and $\mathbf{S}$ are flux functions in the *x, y, z* directions and source terms. ρ, $\mathbf{m}, \mathbf{B}$ and U are the density, momentum, magnetic field and energy density. Where $\rho = \rho_1 + \rho_2$ with ρ_1 and ρ_2 being the densities of first and second plasma components. Using Gauss's law, the integration form of equation (21) can be written as

$$\frac{\partial}{\partial t}\int \mathbf{u}dv + \int (\mathbf{F}n_x + \mathbf{G}n_y + \mathbf{H}n_z)ds = \int \mathbf{S}dv, \tag{22}$$

where dv and ds are the volume and surface element of the control volume and $\mathbf{n}$ is a unit vector normal to the surface of the control volume.

Let us define a matrix $\mathbf{T}$ which rotates the x axis to the direction of $\mathbf{n}$

$$\mathbf{T} = \begin{vmatrix} 1 & & & & \\ & \mathbf{T}_1 & & & \\ & & \mathbf{T}_1 & & \\ & & & 1 & \\ & & & & 1 \end{vmatrix}, \tag{23}$$

with

$$\mathbf{T}_1 = \begin{vmatrix} n_x & n_y & n_z \\ t_{1x} & t_{1y} & t_{1z} \\ t_{2x} & t_{2y} & t_{2z} \end{vmatrix}, \tag{24}$$

then equation (22) is expressed as

$$\frac{\partial}{\partial t}\int \mathbf{u}dv + \int \mathbf{T}^{-1}\mathbf{T}(\mathbf{F}n_x + \mathbf{G}n_y + \mathbf{H}n_z)ds = \int \mathbf{S}dv, \tag{25}$$

where $\mathbf{t}_1$ and $\mathbf{t}_2$ are unit vectors tangent to the surface of the control volume and orthogonal to each other. Since the form of the MHD equations must be unchanged for the rotation of the coordinate system, the relation

$$\mathbf{T}(\mathbf{F}(\mathbf{u})n_x + \mathbf{G}(\mathbf{u})n_y + \mathbf{H}(\mathbf{u})n_z) = \mathbf{F}(\mathbf{T}\mathbf{u}) = \mathbf{F}(\mathbf{u}_n) \tag{26}$$

must hold [15] [16] [19]. Then one can obtain from equation (25)

$$\frac{\partial}{\partial t}\int \mathbf{u}dv + \int \mathbf{T}^{-1}\mathbf{F}(\mathbf{u}_n)ds = \int \mathbf{S}dv. \tag{27}$$

Introducing new dependent variables $\mathbf{u}_1 = (\rho, \mathbf{m}, \mathbf{B}_1, U_1, \rho_2)^T = (\rho, m_x, m_y, m_z, B_x - B_{0x}, B_y - B_{0y}, B_z - B_{0z}, U - (\mathbf{B}_1{\cdot}\mathbf{B}_0)/\beta - B_0^2/(2\beta), \rho_2)^T$, with the conditions $\partial\mathbf{B}_0/\partial t = rot\mathbf{B}_0 = div\mathbf{B}_0$, then the equation for $\mathbf{u}_1$ can be written in the conservaton-law form as

$$\frac{\partial}{\partial t}\int \mathbf{u}_1 dv + \int \mathbf{T}^{-1}\mathbf{F}(\mathbf{u}_{1n}, \mathbf{B}_{0n})ds = \int \mathbf{S}dv, \tag{28}$$

with $\mathbf{u}_{1n} = \mathbf{T}\mathbf{u}_1$, $\mathbf{m}_n = T_1\mathbf{m} = (m_n, m_{t1}, m_{t2})^T$, $\mathbf{B}_n = \mathbf{T}_1\mathbf{B} = (B_n, B_{t1}, B_{t2})^T$, $\mathbf{B}_{1n} = \mathbf{T}_1\mathbf{B}_1 = (B_{1n}, B_{1t1}, B_{1t2})^T$, and $\mathbf{B}_{0n} = \mathbf{T}_1\mathbf{B}_0 = (B_{0n}, B_{0t1}, B_{0t2})^T$. The flux function in the normalized form is written

$$\mathbf{F} = \begin{bmatrix} m_n \\ P + \frac{m_n m_n}{\rho} + \frac{B^2}{2\beta} - \frac{1}{\beta}B_nB_n - \frac{B_0^2}{2\beta} + \frac{1}{\beta}B_{0n}B_{0n} \\ \frac{m_{t1}m_n}{\rho} - \frac{1}{\beta}B_{t1}B_n + \frac{1}{\beta}B_{0t1}B_{0n} \\ \frac{m_{t2}m_n}{\rho} - \frac{1}{\beta}B_{t2}B_n + \frac{1}{\beta}B_{0t2}B_{0n} \\ 0 \\ \frac{m_n}{\rho}B_{t1} - \frac{m_{t1}}{\rho}B_n \\ \frac{m_n}{\rho}B_{t2} - \frac{m_{t2}}{\rho}B_n \\ \frac{m_n}{\rho}(U_1 + \frac{B_1^2}{2\beta} + P) - \frac{B_{1n}}{\beta} \\ \times(\frac{m_n}{\rho}B_{1n} + \frac{m_{t1}}{\rho}B_{1t1} + \frac{m_{t2}}{\rho}B_{1t2}) \\ + \frac{B_{1t1}}{\beta}(\frac{m_n}{\rho}B_{0t1} - \frac{m_{t1}}{\rho}B_{0n}) \\ + \frac{B_{1t2}}{\beta}(\frac{m_n}{\rho}B_{0t2} - \frac{m_{t2}}{\rho}B_{0n}) \\ \frac{\rho_2}{\rho}m_n \end{bmatrix}. \tag{29}$$

In the solar wind-magnetosphere-ionosphere (S-M-I) interaction problem, a dipole field will be adopted as $\mathbf{B}_0$. In the expression of (29), the $\mathbf{B}_0$ terms are added to the second, third, fourth, and last components of F, considering $rot\mathbf{B}_0{\times}\mathbf{B}_0 = 0$ and $\mathbf{m}{\times}\mathbf{B}{\cdot}rot\mathbf{B}_0 = 0$. The variable component of energy density U_1, density ρ, momentum $\mathbf{m}$, and the variable components of magnetic field $\mathbf{B}_1$ are related to pressure P by the equation

$$P = (\gamma - 1)(U_1 - \frac{m^2}{2\rho} - \frac{B_1^2}{2\beta}). \tag{30}$$

Constants in these equations are β and γ, with $\beta = \mu\rho_0 RT_0/B_{00}^2, \gamma$ the polytropic index, μ the magnetic permeability, R the gas constant, ρ_0 the normalization density, B_{00} the normalization field, and T_0 the normalization temperature. Momentum $\mathbf{m}$ and time t are normalized by $\rho_0(RT_0)^{1/2}$ and $L_0/(RT_0)^{1/2}$, with L_0 normalization length.

From equation (28), a discrete formulation of the MHD equations in the FVM style is written for the grid point i in the form

$$\frac{\partial}{\partial t}\mathbf{u}_{1i}V_i + \sum\nolimits_j \mathbf{T}_{ij}^{-1}\underline{\mathbf{F}}_{ij}(\mathbf{u}_{1ni'}, \mathbf{u}_{1ni}, \mathbf{u}_{inj}, \mathbf{u}_{1nj'}, \mathbf{B}_{0nij})S_{ij} = \mathbf{S}V_i, \tag{31}$$

where j denotes the grid points neighboring the grid point i, V_i denotes the volume of the control volume cell which includes the grid point i, $\mathbf{T}_{ij}$ is the rotation matrix at the interfacing surface between i and j, S_{ij} is the surface area of the i and j interface, $\mathbf{u}_{1ni'}$, $\mathbf{u}_{1ni}$, $\mathbf{u}_{1nj}$, and $\mathbf{u}_{1nj'}$ are $\mathbf{u}_{1i'}$, $\mathbf{u}_{1i}$, $\mathbf{u}_{1j}$, and $\mathbf{u}_{1j'}$ rotated by $\mathbf{T}_{ij}$, and $\mathbf{B}_{0nij}$ is $\mathbf{B}_{0n}$ at the i and j interface. Adopting equation (17), the first-order upwind numerical flux $\underline{\mathbf{F}}_{ij}$ for equation (31) is given as

$$\underline{\mathbf{F}}_{ij} = \frac{1}{2}[\mathbf{F}(\mathbf{u}_{1nj}, \mathbf{B}_{0nij}) + \mathbf{F}(\mathbf{u}_{1ni}, \mathbf{B}_{0nij}) - \mathbf{R}_{ij} \mid \mathbf{\Lambda}_{ij} \mid \mathbf{R}_{ij}^{-1}(\mathbf{u}_{1nj} - \mathbf{u}_{1ni})]. \tag{32}$$

Here, the mode synthesis matrix $\mathbf{R}_{ij}$ and the eigenvalue matrix $\mathbf{\Lambda}_{ij}$ are calculated from the following diagonalization process:

$$\mathbf{A}_{ij}\mathbf{R}_{ij} = \mathbf{R}_{ij}\mathbf{\Lambda}_{ij}, \tag{33}$$

$$\mathbf{A}_{ij} = \frac{\partial \mathbf{F}}{\partial \mathbf{u}_{1n}}(\mathbf{u}_{inij}, \mathbf{B}_{0nij}), \tag{34}$$

with $\mathbf{A}_{ij}$ the flux Jacobian matrix of $\mathbf{F}$ at the i and j interface, and $\mathbf{u}_{1nij}$ a symmetric average of $\mathbf{u}_{1nj}$ and $\mathbf{u}_{1ni}$. As seen from equation (33), the mode synthesis matrix consists of the right eigenvectors of the flux Jacobian matrix $\mathbf{r}_{ijk}$, and the diagonal matrix $\mathbf{\Lambda}_{ij}$ consists of eigenvalues λ_{ijk}, $k = 1 \sim 9$.

To get a higher order of accuracy, the MUSCL approach is used changing i and j in equation (32) to L and R, suffixes which indicate variables just on the negative and positive sides of the interface [17]. Adopting equations (18) and (19), then the numerical flux is defined by the following relation:

$$\underline{\mathbf{F}}_{ij} = \frac{1}{2}[\mathbf{F}(\mathbf{u}_{1nR}, \mathbf{B}_{0nij}) + \mathbf{F}(\mathbf{u}_{1nL}, \mathbf{B}_{0nij}) - \mathbf{R}_{RL} \mid \mathbf{\Lambda}_{RL} \mid \mathbf{R}_{RL}^{-1}(\mathbf{u}_{1nR} - \mathbf{u}_{1nL})], \tag{35}$$

with

$$\mathbf{A}_{RL}\mathbf{R}_{RL} = \mathbf{R}_{RL}\mathbf{\Lambda}_{RL}, \tag{36}$$

$$\mathbf{u}_{1nL} = \mathbf{u}_{1ni} + \mathbf{s}_i(\mathbf{1} - \mathbf{s}_i/3)(\mathbf{u}_{1ni} - \mathbf{u}_{1ni'}) + (\mathbf{1} + \mathbf{s}_i/3)(\mathbf{u}_{1nj} - \mathbf{u}_{1ni})/4, \tag{37}$$

$$\mathbf{u}_{1nR} = \mathbf{u}_{1nj} - \mathbf{s}_j(\mathbf{1} - \mathbf{s}_j/3)(\mathbf{u}_{1nj'} - \mathbf{u}_{1nj}) + (\mathbf{1} + \mathbf{s}_j/3)(\mathbf{u}_{1nj} - \mathbf{u}_{1ni})/4, \tag{38}$$

where the diagonal matrices $\mathbf{s}_i$ and $\mathbf{s}_j$ consist of the so-called Van Leer's differentiable limiter. The k-th components of s_i and s_j are calculated from the k-th

components of $\mathbf{u}_{1n}$. Without the suffixes 1 and k, they are written as

$$s_i = \frac{2(u_{nj} - u_{ni})(u_{ni} - u_{ni'}) + \epsilon}{(u_{nj} - u_{ni})^2 + (u_{ni} - u_{ni'})^2 + \epsilon}, \tag{39}$$

$$s_j = \frac{2(u_{nj'} - u_{nj})(u_{nj} - u_{ni}) + \epsilon}{(u_{nj'} - u_{nj})^2 + (u_{nj} - u_{ni})^2 + \epsilon}, \tag{40}$$

with ϵ a small number. Interpolation points i ' and j ' are obtained by extending the line which connects grid points i and j to the neighboring surface of control volumes.

A serious problem in numerical MHD simulations involves the violation of the $div\mathbf{B} = 0$ condition. Not only numerical roundoff errors but also the use of upwind fluxes and a non-Cartesian grid system make it difficult to fulfill the $div\mathbf{B} = 0$ condition automatically. In the present calculation, an extra equation is added to eliminate artificial magnetic monopoles [21]. The variable components of magnetic field $\mathbf{B}_1$ are replaced every several time steps, by a new field $\mathbf{B}_{1c}$ given as

$$\mathbf{B}_{1c} = \mathbf{B}_1 + grad\phi, \tag{41}$$

$$\bigtriangledown^2 \phi = -div\mathbf{B}_1. \tag{42}$$

To solve equation (41), the conjugate residual (CR) method is applied.

4 Eigenvalues and Eigenvectors for the 9-Component MHD Equations

In this section, eigenvalues and eigenvectors are shown for the 9-component MHD equations. For the Jacobian matrix of flux fuction (29), eigenvalues λ_k, $k = 1 \sim 9$ are [14]

$$\lambda_1 = m'_n, \tag{43}$$

$$\lambda_{2,3} = m'_n \pm |B'_n|, \tag{44}$$

$$\lambda_{4,5} = m'_n \pm V_f, \tag{45}$$

$$\lambda_{6,7} = m'_n \pm V_s, \tag{46}$$

$$\lambda_8 = 0, \tag{47}$$

$$\lambda_9 = m'_n, \tag{48}$$

where

$$V_f^2, V_s^2 = \frac{1}{2}[C_0 + B'^2 \pm \{(C_0 + B'^2)^2 - 4C_0 {B'_n}^2\}^{1/2}], \tag{49}$$

$$C_0 = \gamma P/\rho, \tag{50}$$

with the notation $\mathbf{u}'_n = (\rho, \mathbf{m}'_n, \mathbf{B}'_n, U, \rho_2)^T = (\rho, m_n/\rho, m_{t1}/\rho, m_{t2}/\rho, B_n/\sqrt{\beta\rho}, B_{t1}/\sqrt{\beta\rho}, B_{t2}/\sqrt{\beta\rho}, U, \rho_2)^T$. In the expresion of eigenvalues, $\sqrt{C_0}, |B'_n|, V_f$ and V_s correspond to sound, Alfen, fast and slow velocities, respectively. In addition to one entropy, two Alfven, two fast and two slow waves of normal MHD

equations, there appears one more entropy wave in the 9-component MHD equations. Calculations of eigenvectors must be done with special care avoiding the degeneration of eigenvectors when wave propagations become perpendicular or parallel to the magnetic field. The right eigenvectors $\mathbf{r}_k$ which correspond to λ_k are [19]

$$\mathbf{r}_1 = \begin{bmatrix} 1 \\ m'_n \\ m'_{t1} \\ m'_{t2} \\ 0 \\ 0 \\ 0 \\ 0.5{\cdot}m'^2 \\ 0 \end{bmatrix}, \tag{51}$$

$$\mathbf{r}_{2,3} = \begin{bmatrix} 0 \\ 0 \\ \mp B''_{t2}{\cdot}sgn(B'_n) \\ \pm B''_{t1}{\cdot}sgn(B_n) \\ 0 \\ B''_{t2}\sqrt{\frac{\beta}{\rho}} \\ -B''_{t1}\sqrt{\frac{\beta}{\rho}} \\ \mp(B''_{t2}m'_{t1} - B''_{t1}m'_{t2}){\cdot}sgn(B'_n) + (B''_{t2}B'_{1\,t1} - B''_{t1}B'_{1\,t2}) \\ 0 \end{bmatrix}, \tag{52}$$

$$\mathbf{r}_{4,5} = \begin{bmatrix} a_f \\ a_f(m'_n \pm V_f) \\ a_f m'_{t1} \mp a_s B''_{t1} V_f B'_n \\ a_f m'_{t2} \mp a_s B''_{t2} V_f B'_n \\ 0 \\ a_s B''_{t1} V_f^2 \sqrt{\frac{\beta}{\rho}} \\ a_s B''_{t2} V_f^2 \sqrt{\frac{\beta}{\rho}} \\ a_f{\cdot}0.5{\cdot}m'^2 + a_f V_f^2/(\gamma-1) \pm a_f V_f m'_n \\ \mp a_s V_f (B''_{t1} m'_{t1} + B''_{t2} m'_{t2}) B'_n \\ +a_f(-1)/(\gamma-1)(V_f^2 - C_0) \\ +a_f(V_f^2 - C_0)(B''_{t1}B''_{1\,t1} + B''_{t2}B''_{1\,t2}) \\ /(B''^{\,2}_{t1} + B''^{\,2}_{t2}) \\ a_f \rho_2/\rho \end{bmatrix}, \tag{53}$$

$$\mathbf{r}_{6,7} = \begin{bmatrix} a_s \\ a_s(m'_n \pm V_s) \\ a_s m'_{t1} \pm a_f B''_{t1}\sqrt{C_0}/V_f \cdot sgn(B'_n) \\ a_s m'_{t2} \pm a_f B''_{t2}\sqrt{C_0}/V_f \cdot sgn(B'_n) \\ 0 \\ -a_f B''_{t1}\sqrt{\frac{\beta}{\rho}}C_0/V_f^2 \\ -a_f B''_{t2}\sqrt{\frac{\beta}{\rho}}C_0/V_f^2 \\ a_s \cdot 0.5 \cdot m'^2 + a_s V_s^2/(\gamma-1) \pm a_s V_s m'_n \\ \pm a_f(B''_{t1}m'_{t1} + B''_{t2}m'_{t2}) \\ \times\sqrt{C_0}/V_f \cdot sgn(B'_n) + a_s(-1)/(\gamma-1)(V_s^2 - C_0) \\ +a_s(V_s^2 - C_0)(B''_{t1}B''_{1\,t1} + B''_{t2}B''_{1\,t2}) \\ /(B''^{\,2}_{t1} + B''^{\,2}_{t2}) \\ a_s\rho_2/\rho \end{bmatrix}, \tag{54}$$

$$\mathbf{r}_8 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{55}$$

$$\mathbf{r}_9 = \begin{bmatrix} 1 \\ m'_n \\ m'_{t1} \\ m'_{t2} \\ 0 \\ 0 \\ 0 \\ 0.5 \cdot m'^2 \\ 1 \end{bmatrix}, \tag{56}$$

where

$$B''_{t1} = (B'_{t1} + \epsilon)/(B'^2_{t1} + B'^2_{t2} + 2\epsilon^2)^{1/2}, \tag{57}$$

$$B''_{t2} = (B'_{t2} + \epsilon)/(B'^2_{t1} + B'^2_{t2} + 2\epsilon^2)^{1/2}, \tag{58}$$

$$B''_{1\,t1} = (B'_{1\,t1} + \epsilon)/(B'^2_{t1} + B'^2_{t2} + 2\epsilon^2)^{1/2}, \tag{59}$$

$$B''_{1\,t2} = (B'_{1\,t2} + \epsilon)/(B'^2_{t1} + B'^2_{t2} + 2\epsilon^2)^{1/2}, \tag{60}$$

$$a_f = (V_f^2 - B'^2_n)^{1/2}/(V_f^2 - V_s^2)^{1/2}, \tag{61}$$

$$a_s = (V_f^2 - C_0)^{1/2}/(V_f^2 - V_s^2)^{1/2} V_f, \tag{62}$$

and ϵ is a small number.

5 Source Terms and Boundary Conditions

The selection of source terms and boundary conditions depends on the kind of problem treated by the MHD simulation. Typical source terms considered in space science are ion production and loss, gravitational acceleration and aeronomic friction. They are written as

$$\mathbf{F} = \begin{vmatrix} q_1 + q_2 - L_1 - L_2 \\ -\nu\mathbf{m} - \rho\mathbf{g} \\ \mathbf{0} \\ -\mathbf{m}/\rho\cdot(\nu\mathbf{m} + \rho\mathbf{g}) + T_q(q_1 + q_2)/(\gamma - 1) - T_L(L_1 + L_2)/(\gamma - 1) \\ q_2 - L_2 \end{vmatrix}, \quad (63)$$

where q_i and L_i $(i = 1, 2)$ are ion production and loss terms for i-th ion species, T_q is the temperature of ions when they are produced, T_L is the temperature of ions when they are lost. ν and $\mathbf{g}$ are ion-neutral collision frequency and gravitational acceleration. Production and loss rates q_i and L_i $(i = 1, 2)$ are normalized by ρ_0 and $L_0/(RT_0)^{1/2}$.

A typical boundary condition on the outer boundary is to give a plasma flow on the upstream side and a zero gradient condition on the downstream side. Commonly used boundary conditions on the inner boundary are ion chemical equilibrium, given plasma velocity, or zero gradient condition. In the case of S-M-I coupling, the plasma velocity perpendicular to the ambient magnetic field is decided from the field aligned current (FAC) flowing into the ionosphere. Assuming a spherical ionosphere at r=1 R_e, these processes are simulated on the inner boundary and on the ionosphere from

$$\nabla\cdot\sigma\nabla\phi_I = G_m(rot\mathbf{B}_1\cdot\mathbf{n}_b) = J_\parallel, \quad (64)$$

$$\sigma = \sigma_{EUV} + \sigma_{Diff}(P, \rho) + \sigma_J(J_\parallel), \quad (65)$$

$$\phi_m = \phi_I - f_1(J_\parallel)|J_\parallel|, \quad (66)$$

$$\mathbf{m} - (\mathbf{m}\cdot\mathbf{n}_b)\mathbf{n}_b = -\rho\nabla\phi_m\times\mathbf{B}/B^2, \quad (67)$$

where σ is the ionospheric conductivity tenser, ϕ_I is the ionospheric potential, ϕ_m is the magnetospheric potential, $J_\parallel$ is the FAC, $\mathbf{n}_b$ is a unit vector along $\mathbf{B}$, and G_m is a geometrical factor associated with the mapping along field lines from r=3 R_e to r=1 R_e. σ_{EUV}, σ_{Diff}, and σ_J are the ionospheric conductivities due to the solar EUV, the diffuse auroral precipitation modeled by the pressure and temperature, and discreet precipitation modeled by the upward FAC [5]. Parallel potential is introduced through f_1 that is constant at the upward FAC and 0 at the downward FAC.

6 Parallel Computation in a Spherical Geometry

Recently, the most powerful super-computers, such as the Fujitsu VPP and NEC SX, adopt the vector-parallel architecture with a distributed memory system.

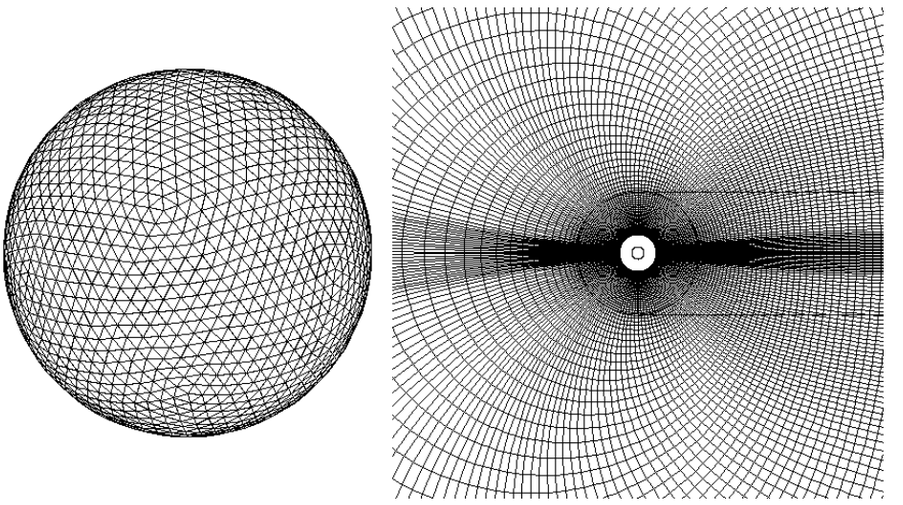


Fig. 1. Grid structure for the 3-D FVM calculation

Efficient utilization of these vector-parallel super-computers is essential for the future study of space MHD simulations. In parallel computations with a distributed memory system, it is desirable to set a one-dimensional structuring axis in the three-dimensional space. In a spherical geometry that is important for space science, this "parallel" axis is chosen to be the radial direction. An unstructured grid can then be generated on spherical surfaces which construct the remaining two-dimensional space.

In the construction process of the grid system, it is desirable that two-dimensional spherical surfaces are covered by control volumes of similar size, because the integration time step is restricted by the smallest control volume. An example for such kind of grid system is shown in Fig. 1. The left panel in Fig. 1 shows the grid structure on the spherical inner boundary, while the right panel shows how to construct a 3-D grid structure by extending the position of spherically allocated grids outward from the inner boundary.

In parallel computations on the distributed memory system of vector-parallel computers, it is important to identify the difference between distributed and redundant data areas. The parallel axis is used to distribute dependent variables to processors and to define the overlap data areas. In the finite volume TVD scheme, the most serious load in calculations comes from those for eigenvalues (43-50), eigenvectors (51-62), limiters (39,40) and numerical flux (29). These calculations are done sequentially in subroutines on a two-dimensional redundant data area, after copying 3-D dependent variables from a distributed data area to two-dimensionl redundant data areas. Using this method means that we do not have to rewrite subroutines with parallel programming constructs. The parallelization then occurs only in the main program which calls these subroutines. In the calculation of numerical flux, the dependent variables on the neighboring

grid point of the calculation point must be referred to. To enable this referring in the parallel computation, the overlap data areas are used with the data transmission. Overlap data must be synchronized to data in the neighboring processors before the subroutines begin concurrent operation. After the main calculations, the final correction of the dependent variables is done in the main program, on the distributed data area. Finally, two processors calculate the inner and outer boundary conditions.

7 Numerical Example 1 (Heliospheric Structure)

The pressure difference between the solar corona and interstellar space drives the ionized solar atmosphere outward, despite the restraining influence of solar gravity. The solar wind thus generated interacts with the very local interstellar medium (VLISM) at some large distance from the sun. Here, the volume of the space created by the solar wind is called the heliosphere [8] [10] [11] [22] [23]. In this section, results are shown for the MHD simulation of the heliosphere.

At a distance where the local ram pressure of the solar wind becomes comparable to the external VLISM pressure, the solar wind shocks to form the termination shock (TS), which is a strong shock with a compression ratio 4. Then outside the TS, the shocked subsonic solar wind flows to the downstream direction of a uniform interstellar flow surrounding the heliosphere. This region constrained by the VLISM and filled with shocked solar wind plasma is called as heliosheath (HS). It is bounded inside by the TS and outside by a tangential discontinuity between the heliosphere and the VLISM called the heliopause (HP). Some observational evidences suggests that the interstellar wind is also supersonic. The supersonic interstellar flow recognizes the shocked solar wind plasma in the HS as an obstacle and forms a bow shock (BS), which is a magnetoacoustic shock. The shocked interstellar wind inside the BS contacts the shocked solar wind through the HP.

Since the formation process of the heliosphere generates many shocks and discontinuities, it offers a good problem to test the shock-capturing capability of the TVD scheme. The outer and inner boundaries for the calculation are set at 1000 AU and 50 AU. In this section, the interstellar plasma and the solar wind plasma are assigned to ρ_1 and ρ_2, respectively. Consequently, ρ_1 is zero at the inner boundary and ρ_2 is zero at the upstream boundary. On the inner boundary, a supersonic solar wind is adopted. The solar wind speed and density at 1 AU are assumed to be 400 km/sec and 5 cm^{-3} respectively, and the strength of the toroidal interplanetary magnetic field (IMF) here is assumed to be $2.8 * cos(\theta)$ nT with θ the heliolatitude. Toward the outer boundary, the solar wind maintains a constant velocity while its density and magnetic field fall with heliocentric distance r as r^{-2} and r^{-1}. The solar wind temperature at the inner boundary is assumed to be 10^4 K. The speed, density and temperature of the interstellar medium are assumed to be 25 km/sec, 0.1 cm^{-3}, and 10^4 K. The direction of the interstellar flow and magnetic field are assumed to be parallel to the ecliptic plane (toward -x) and to the solar rotational axis (toward +z). The strength

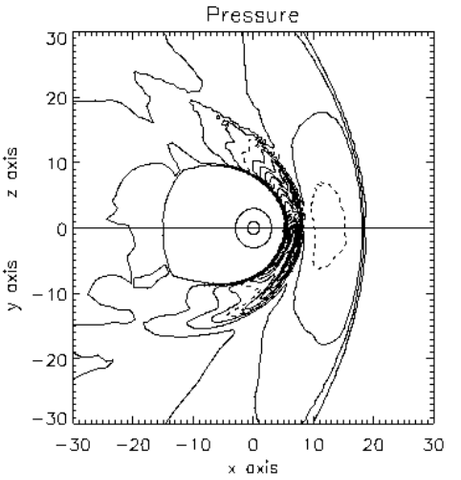


Fig. 2. Pressure distribution in the heliosphere

of interstellar magnetic field is 0.15 nT. $\mathbf{S}$ and $\mathbf{B}_0$ are not considered in this problem.

Figure 2 shows the normalized equipressure (P) contour on the polar (upper half) and ecliptic (lower half) planes. The interstellar wind is from the right. The normalization value for P and contour spacing are 0.0144 pPa and 0.7, respectively. From the pressure distribution, the major structures of the heliosphere, the TS with Mach disk, the HP, and the BS are clearly visible as discontinuities. These high-quality resolutions of discontinuities are due to the excellent shock-capturing property of the TVD scheme.

At the BS the kinetic energy of the interstellar wind is converted to thermal and magnetic energies. Downstream of the BS, consequently, gas pressure dominates over the kinetic pressure. At the HP, increased gas and magnetic pressure are supported by the HS plasma pressure which is maintained by a supply of shocked solar wind pressure from the TS. The highest pressure in the HS appears in the nose region because it must finally balance with the dynamic pressure of the interstellar wind.

The enhanced HS pressure around the nose region accelerates the shocked solar wind plasma toward the heliotail (HT). At the flank of the heliosphere, the oblique TS also helps the HS flow direct downtail based on the principle that at an oblique shock the downstream flow is always deflected away from the shock normal. On the contrary, the tailside TS consists of right-angle shock, because the post-shock flow can direct downtail as it is. Consequently, tangential discontinuity develops in the downstream HT, to separate fast HT flow that continues from the flank HS from a slow HT flow that exits directly from the tailside TS. This is a basic mechanism to form a bullet-shaped TS (Mach disk) on the downstream side [11].

8 Numerical Example 2 (Solar Wind-Venus Interaction)

Since the internal magnetic field of Venus is negligibly small, the solar wind makes a direct contact with the ionosphere. When the interaction processes between the solar wind and planetary ionospheres are studied, at least two plasma components must be considered [9] [24] [25]. In the case of venusian ionosphere, the primary component of ionospheric plasma is O^+ ions, whereas the primary component of the solar wind is H^+ ions. So one must consider two plasma components to distinguish the ionospheric plasma from the solar wind plasma. In this section, therefore, the solar wind plasma and ionospheric plasma are assigned to ρ_1 and ρ_2, respectively. In this problem, source terms $\mathbf{S}$ are essentially important, because the high-density low-temperature ionospheric plasma that supports the impinging solar wind stratifies gravitationally on the balance of ion production and loss, neutral drag, and the gravitational acceleration. For the calculation of ρ_2, L_2, and ν, a stratified atmosphere composed of O and CO_2 is assumed around the planet. q_1, L_1, and $\mathbf{B}_0$ are set to zero in this section.

The inner and outer boundaries of the calculation region are at 1 R_p and 10 R_p, with R_p the planetary radius. On the outer boundary, the solar wind flow is given on the upstream ($+x$) side while the zero gradient condition is adopted on the downstream ($-x$) side. Where the IMF is assumed to be parallel to the y-axis. Near the inner boundary, the ion-neutral collision and ion chemical processes become dominant. Therefore, the ion chemical equilibrium and zero plasma velocity conditions are adopted on the inner boundary. Since the scale sizes of the ionosphere and the solar wind are quite different from each other, the grid points must be allocated so as to be dense in the ionosphere and coarse in the solar wind.

Figure 3 shows the result for the distributions of O^+ (left) and total (=$H^+ + O^+$, right) ion densities. The solar wind is from the left. The left and right panels in Fig. 3 show contours of $log(\rho_2/\rho_{sw})$ and $\log(\rho/\rho_{sw})$, respectively. Where ρ_{sw} is the solar wind density. The contour spacing is 0.2 and the dashed contours

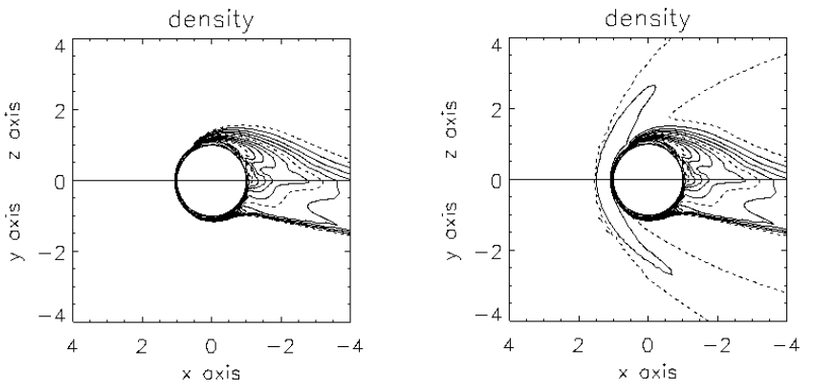


Fig. 3. O^+ density (left) and H^++O^+ density (right) around the planet

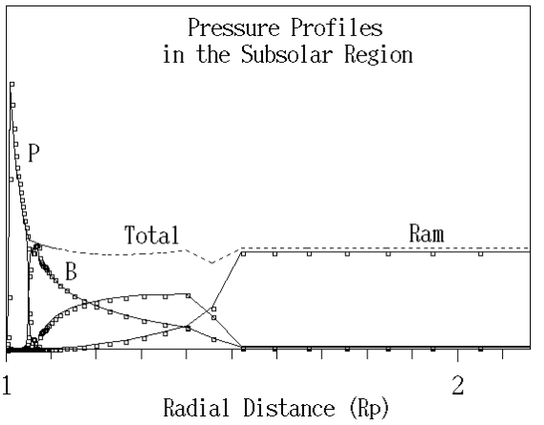


Fig. 4. Pressure balance along the sun-planet line

are used at every 1.0. The minimum contour value for a dashed contour is 0.0. The upper and lower halves of the two panels show sun-planet meridian planes and equatorial planes defined by the direction of the IMF. The solid circles show the size of the planet. An excellent capturing of shocks and discontinuities is seen in Fig. 3. Results of the calculation show the formation of the BS, magnetic barrier and the ionopause in the dayside region. At the ionopause, the primary ion species change from H^+ on the high-altitude side to O^+ on the low-altitude side. In the nightside region, the ionospheric structure shows rather complex features. A part of O^+ ions penetrates into the magnetotail which results from the draping process of the IMF. Then, the penetrating O^+ ions tend to gather toward the central part of the magnetotail and form a high-density region. In these results, the solar wind and ionospheric plasmas are distinguished clearly by the 9-component MHD equations.

Figure 4 shows altitude distributions of fluid pressure (P), magnetic pressure (B) and plasma dynamic pressure (Ram) at the subsolar point. Where small rectangles show positions of radial grid points. The pressures are normalized by the solar wind pressure. In the upstream solar wind, plasma kinetic energy dominates both of fluid pressure and magnetic pressure. At the BS, plasma kinetic energy is converted to plasma thermal energy. As a result, the fluid pressure becomes dominant after passing the BS. Approaching the ionopause, the magnetic pressure increases while the plasma pressure decreases, due to the formation of the magnetic barrier. At the ionopause, the magnetic barrier is supported by the fluid pressure of cold ionospheric plasma. This pressure of cold ionospheric plasma is maintained by the photoionization and ion chemical processes in the planetary upper atmosphere. The plasma pressure on the bottom side of the ionosphere is supported by the neutral atmosphere through ion-neutral collisions.

9 Numerical Example 3 (Substorm and Space Weather)

A goal of the space weather effort is to increase our understanding of the S-M-I coupling system. In recent years, the global MHD simulation has become increasingly successful at constructing and predicting the behavior of the S-M-I system [1] [5] [6] [12]. It gives a theoretical foundation for the complex behavior of the S-M-I system that is controlled by the coupling process between different regions.

In the solar wind interaction with the magnetosphere, energy and momentum are transferred from the solar wind to the magnetosphere through non-ideal MHD processes, to generate magnetospheric plasma convection [4]. Figure 5 schematically shows the construction of convection system. In the magnetosphere, the large-scale transportation of plasma is equivalent to a global electric field. The process driving the magnetospheric convection is at the same time the process generating the FACs, because the magnetospheric perpendicular stress must be transmitted to the polar ionosphere so as the ionospheric convection to follow the magnetospheric convection [3] [26]. In the current circuit connecting the magnetospheric dynamo and the conducting ionosphere, the $\mathbf{J}\times\mathbf{B}$ force in the ionosphere acts to accelerate the ionospheric convection against atmospheric friction. As a counter part of this energy dissipation in the ionosphere, FACs must be powered through the dynamo driven by the energy conversion in the convection system. Therefore, acting as a load for the magnetospheric convection in the M-I coupling system, the ionosphere controls the intensity of FAC.

In this section, we investigate the M-I processes that maintain the self-consistency in the convection system, including the generation mechanism of the M-I current systems, the ionospheric control of the magnetospheric configuration, and possible extension of convection status to the substorm. The FAC and plasma convection play a central role in the M-I coupling, while the state of energy source for these current systems depends on the solar wind-magnetospheric interaction. To improve our understanding of this problem, therefore, a self-consistent treatment is required for the coupling effects between three different regions, namely the solar wind, the magnetosphere and the ionosphere.

In the numerical study of the S-M-I coupling process, numerical errors in the low-β region near the ionosphere should be reduced. For this purpose, the MHD calculation is reconstructed as shown in equation (29) to suppress the direct inclusion of the potential component of the magnetic field as dependent variables. From these situations, a dipole magnetic field is assumed as $\mathbf{B}_0$. $\mathbf{S}$ and ρ_2 are not considered in this section. The outer and inner boundaries for the simulation are at 200 R_e and 3 R_e. A uniform solar wind with a speed of 350 km/sec and an IMF magnitude of 5 nT is assumed at the upstream boundary and zero gradients are assumed at the downstream boundary. Dependent variables are projected along the field line from the inner boundary (3.0 R_e) to the ionosphere. In the ionosphere, equations (64) and (65) are solved to match the divergence of the Pedersen and Hall currents with the FAC.

Figure 6 shows the response of the magnetosphere to the southward turning of the IMF. The color figure shows the pressure distribution in the noon-midnight

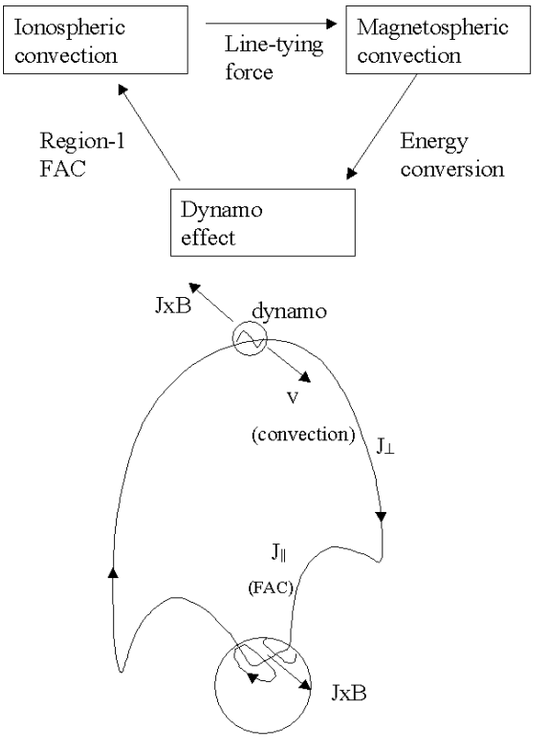


Fig. 5. Magnetospheric convection

meridian plane of the magnetosphere at three times. Where P is normalized by the solar wind P. The bottom row illustrates the initial magnetospheric configuration for the northward IMF. At this time (7.7 minutes after the southward turning of the IMF), a thick and low-pressure plasma sheet is observed. The flow structure at this time (not shown) indicates that x line is situated beyond x =-60 R_e, which is the remnant of merging cell structure under the northward IMF condition [4] and called the distant neutral line. The growth phase shown in the second row (59 minutes after the southward turning of the IMF) is characterized by erosion of the dayside magnetosphere, thinning of the plasma sheet, and an increase in the flaring angle. The tail-like configuration of the plasma sheet during the growth phase is the consequence of an enhanced convection.

The substorm onset occurs as an abrupt change of the magnetospheric configuration in the near-earth tail. The top row in Fig. 6 (72.6 minutes after the southward turning of the IMF) shows the pressure distribution after the onset illustrating the appearance of the high-pressure region in the inner magnetosphere and the formation of the NENL in the midtail. Figure 6 also shows pressure and V_x distributions along the -x axis in the near-earth and midtail regions before and after the onset. Where V_x is normalized by the solar-wind sound velocity. After t=70 min, a sudden change of pressure profile is seen to start just like

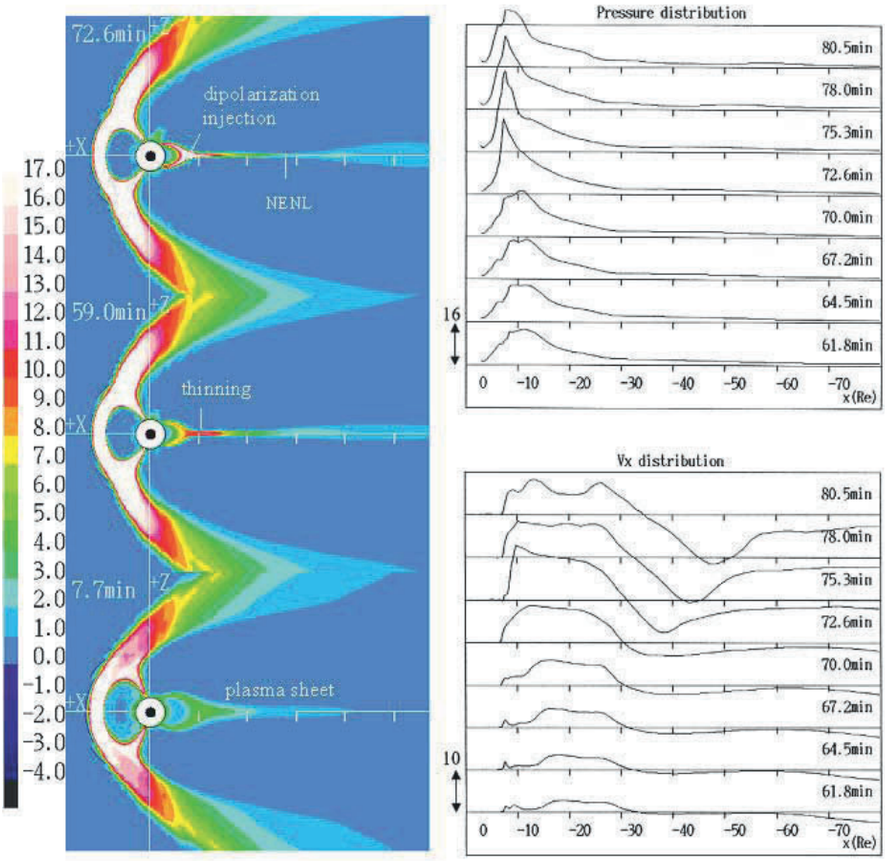


Fig. 6. Substorm sequence obtained from the MHD simulation

a transition from one state to another [5]. Before the onset (t<70 min), the strongest $-\nabla P$ force acts in the region between x=-10 and -20 R_e. As a result, earthward convection is obstructed at x=-14 R_e. In addition, a gradual formation of NENL is seen at x=-33 R_e before the onset. After the onset (t>70 min), the peak position in the pressure distribution shows a rapid inward movement. The pressure peek abruptly moves further inward to x=-8 R_e. At the same time, the convection flow intrudes into the inner magnetosphere inside x=-10 R_e increasing in magnitude. Through these transition processes, a new stress balance is achieved in the near-earth plasma sheet in which recovered magnetic tension is balanced by newly established pressure inside x=-10 R_e. This pressure change is, in turn, a result of energy conversion from magnetic energy to internal energy caused by the pumping effect of convection associated with the recovery of magnetic tension. The fastest earthward flow in the plasma sheet appears after

about 5 minutes from the onset. Then, tailward flow increases its speed. After t=75.3 min, the NENL begins to gradually retreat downtail.

During substorms, the ionospheric conductance enhances to a large extent due to precipitating particles which carry enhanced FACs. As a result, the coupling rate between the magnetosphere and the ionosphere becomes stronger, and the magnetosphere comes to hold a heavier load. However, the role of a variable M-I coupling in substorm onset is not clear at the present time [5].

References

1. T. Ogino: J. Geophys. Res. **91**, 6791 (1986)
2. T. Tanaka: J. Geophys. Res. **98**, 17,251 (1993)
3. T. Tanaka: J. Geophys. Res. **100**, 12,057 (1995)
4. T. Tanaka: J. Geophys. Res. **104**, 14,683 (1999)
5. T. Tanaka: J. Geophys. Res. **105**, 21,081 (2000)
6. J. A. Fedder, J. G. Lyon: J. Geophys. Res. **100**, 3623 (1995)
7. S. Cable, R. S. Steinolfson: J. Geophys. Res. **100**, 21,645 (1995)
8. H. Washimi, T. Tanaka: Space Sci. Rev. **78**, 85 (1996)
9. T. Tanaka, K. Murawski: J. Geophys. Res. **102**, 19,805 (1997)
10. T. J. Linde, T. I. Gombosi, P. L. Roe, G. Powell, D. L. DeZeeuw: J. Geophys. Res. **103**, 1889 (1998)
11. T. Tanaka, H. Washimi: J. Geophys. Res. **104**, 12,605 (1999)
12. G. L. Siscoe, N. U. Crooker, G. M. Erickson, B. U. O. Sonnerup, K. D. Siebert, D. R. Weimer, W. W. White, N. C. Maynard: 'Global geometry of magnetospheric currents'. In: *Magnetosperic Current System*, ed. by S. I. Ohtani (AGU, Washington D.C. 2000) pp. 41-52
13. M. Vinokur: J. Comput. Phys. **82**, 1 (1989)
14. M. Brio, C. C. Wu: J. Comput. Phys. **75**, 400 (1988)
15. T. Tanaka: Comp. Fluid Dyn. J. **1**, 14 (1992)
16. T. Tanaka: SIMPO Newsl. **3**, 31 (1993)
17. B. van Leer: J. Comput. Phys. **32**, 101 (1979)
18. H. C. Yee, G. H. Klopfer, J. L. Montagne: J. Comput. Phys. **88**, 31 (1990)
19. T. Tanaka: J. Comput. Phys. **111**, 382 (1994)
20. K. Murawski, T. Tanaka: Astrophys. Space Sci. **25**4, 187 (1997)
21. J. Brackbill, D. Barnes: J. Comput. Phys. **35**, 426 (1980)
22. E. N. Parker: *Interplanetary Dynamical Processes* (Interscience, New York 1963)
23. H. Washimi, T. Tanaka: Adv. Space Res. **23(3)**, 551 (1999)
24. T. Tanaka: Earth Planets Space **50**, 259 (1998)
25. T. Tanaka: Adv. Space Res. **26(10)**, 1577 (2000)
26. T. Tanaka: 'Generation mechanism of the field-aligned current system deduced from a 3-D MHD simulation of the solar wind-magntosphere-ionosphere coupling'. In: *Magnetospheric Research with Advanced Techniques*. ed. by R. L. Xu, A. T. Y. Lui (Pergamon, 1998) pp. 133-142

Global Magnetohydrodynamic Simulation Using High Performance FORTRAN on Parallel Computers

Tatsuki Ogino

Solar-Terrestrial Environment Laboratory, Nagoya University 3-13 Honohara, Toyokawa, Aichi 442-8507, Japan

Abstract. High Performance Fortran (HPF) is one of modern and common techniques to achieve high performance parallel computation. We have translated a 3-dimensional magnetohydrodynamic (MHD) simulation code of the Earth's magnetosphere from VPP Fortran to HPF/JA on the Fujitsu VPP5000/56 vector-parallel supercomputer and the MHD code was fully vectorized and fully parallelized in VPP Fortran. The entire performance and capability of the HPF MHD code could be shown to be almost comparable to that of VPP Fortran. A 3-dimensional global MHD simulation of the earth's magnetosphere was performed at a speed of over 400 Gflops with an efficiency of 76.5 VPP5000/56 in vector and parallel computation that permitted comparison with catalog values. We have concluded that fluid and MHD codes that are fully vectorized and fully parallelized in VPP Fortran can be translated with relative ease to HPF/JA, and a code in HPF/JA may be expected to perform comparably to the same code written in VPP Fortran.

1 Introduction

If you want to apply modern and common techniques to achieve efficient parallel computation, there are generally two choices of High Performance Fortran (HPF) and Message Passing Interface (MPI). In the two methods, the advantage to use HPF is in less efforts to translate the program. We describe how one can do it for a MHD code and discuss related questions and practical applicable to achieve high efficiency. We have been able to execute a 3-dimensional global magnetohydrodynamic (MHD) simulation of the interaction between the solar wind and the earth's magnetosphere in order to study the structure and dynamics of the magnetosphere using a variety of computers to run a fully vectorized MHD code [1]-[5]. Among the computers we have used are: CRAY Y-MP, Fujitsu VP-2600, Hitachi S820, and NEC SX-3. This flexibility of using many different computers allowed us to work together with many scientists in performing our computer simulations. However, as vector parallel and massively parallel supercomputers have come into the simulation community, a number of different approaches are now being used [6]-[8]. Many scientists carrying out simulations have lost their common language and are being forced to work with new dialects.

We began to use the vector parallel supercomputer, Fujitsu VPP500, in the Computer Center of Nagoya University 1995 and have succeeded in rewriting our 3-dimensional global MHD simulation code in VPP Fortran allowing us to

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 296–314, 2003.

achieve a high performance of over 17 Gflops [8],[9]. Moreover, we have used a new supercomputer, the Fujitsu VPP5000/56, since December 1999 to achieve an even higher performance of over 400 Gflops in VPP Fortran [8],[9],[10]. However, the MHD code in VPP Fortran cannot achieve such a high performance using other supercomputers such as Hitachi SR8000 and NEC SX-5. Thus we face a difficult problem in collaborating in with other scientists in computer simulations even at universities within Japan. We very much hope to recover the advantage of having a common language in the supercomputer world. Recent candidates for this common language appear to be High Performance Fortran (HPF) and Message Passing Interface (MPI). We look forward to the time when we can use these compilers in supercomputers.

In 1999 it was learned that some engineers in Japanese supercomputer companies had tried to develop an extended version of HPF created by the Japanese HPF Association (JAHPF) [11] for use in the new HPF compiler in the near future. This information was of great interest because the performance of the original HPF was questionable and we want to test its performance independently. Since June 2000, we have had the opportunity to use HPF/JA (an extended version of HPF) with a supercomputer of the vector-parallel type, Fujitsu VPP5000/56 [11],[12] and began to translate our 3-dimensional MHD simulation code for the Earth's magnetosphere from VPP Fortran to HPF/JA [4],[9],[12]. The MHD code was fully vectorized and fully parallelized in VPP Fortran [6],[8]. We successfully rewrote the code from VPP Fortran to HPF/JA in three weeks, and an additional two weeks were required to perform a final verification of the results of the calculations. The entire performance and capability of the HPF MHD code are shown to be almost comparable to those of the VPP Fortran code in typical simulations using the Fujitsu VPP5000/56 system.

2 From VPP Fortran to HPF/JA

We requested the Computer Center of Nagoya University to begin using HPF in their supercomputer as soon as possible. In June 2000, stimulated by a lecture on HPF by a Fujitsu engineer, we began to use HPF/JA on the Fujitsu VPP5000 [11],[12]. We understood that the use of HPF was not very successful in the USA and Europe, and that users would rather abandon the usage of HPF due to its difficulty in achieving a high performance. However, we have high expectations of HPF/JA because supercomputer companies in Japan have had valuable experiences with compilers in operating vector-parallel machines [8],[9]. We have decided to translate our MHD code from VPP Fortran to HPF if we can reach 50performance of VPP Fortran using HPF. We have experience in fully vectorizing and parallelizing several test programs and 3-dimensional global MHD codes for the earth's magnetosphere using VPP Fortran. Based on that experience, we believe that we can obtain good performance by translating the MHD codes from VPP Fortran to HPF/JA. The main points of our approach are summarized as follows,

(1) We can use the same domain decomposition (z-direction in 3-dimension) as in VPP Fortran, and overlap data at the boundary of distributed data which were handled by "sleeve" in VPP Fortran can be replaced by "shadow" and "reflect" in HPF/JA.
(2) Parallelizing formalisms for decomposition are performed by "independent" sentence as is done in VPP Fortran.
(3) Lump transmission of distributed data is done by "asynchronous" sentence.
(4) "Asynchronous" sentence can be used for lump transmission between the distributed data with different directions of decomposition.
(5) Unnecessary communication, which was a problem in HPF, can be removed by the instructions of "independent new ( )" and "on home local ( )".
(6) Maximum and minimum can be calculated in parallel by "reduction".

Fortran programs using VPP Fortran can be directly rewritten using HPF/JA owing to (1)-(3) keeping their original styles. If unnecessary communication could be completely stopped by (5), the efficiency of parallelization could be greatly improved. We can use the instructions to stop unnecessary communication if the calculation results don't change with insertion of the instructions. Generally it is very difficult for the parallel compiler to know whether the communication is needed or not in advance because variables might be rewritten at any time. Therefore unnecessary communication that users are not aware of frequently happens reducing the efficiency of the parallelization. That is, the compiler just compiles programs considering all cases; it cannot provide a good solution if there exists any uncertain portion, and therefore unexpected and unnecessary communication very often occurs in the execution of a program.

We can succeed in parallelization of the MHD program even in the worst case because the high speed alternation in the directions of decomposition by (4) can solve almost all the difficulties of parallelization in the MHD program, e.g. in the treatment of boundary conditions. We were able to fully parallelize our MHD code by using the function of alternative decomposition when we used Matsusita Electric Co. ADETRAN compiler. We obtained an efficiency of parallelization over 85computer with 256 processors. Moreover, we could execute parallel computation in maximum, minimum and summation by using a "reduction" sentence in HPF/JA. Because we had high expectations for translating programs from VPP Fortran to HPF/JA, we soon began to rewrite several test programs as well as the 3-dimensional MHD code.

3 Practical Aspects of Translation from VPP Fortran to HPF/JA

The 3-dimensional global MHD simulation code for the interaction between the solar wind and the earth's magnetosphere solves the MHD and Maxwell's equations in 3-dimensional Cartesian coordinates (x,y,z) as an initial and boundary value problem using a modified leap-frog method [4]. The normalized MHD

equations used in the simulation are written as follows:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\mathbf{v}\rho) + D\nabla^2 \rho$$

$$\frac{\partial \mathbf{v}}{\partial t} = -(\mathbf{v} \cdot \nabla)\mathbf{v} - \frac{1}{\rho}\nabla p + \frac{1}{\rho}\mathbf{J} \times \mathbf{B} + \mathbf{g} + \frac{1}{\rho}\mathbf{\Phi}$$

$$\frac{\partial p}{\partial t} = -(\mathbf{v} \cdot \nabla)p - \gamma p \nabla \cdot \mathbf{v} + D_p \nabla^2 p$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}$$

$$\mathbf{J} = \nabla \times (\mathbf{B} - \mathbf{B}_d)$$

where ρ is the plasma density, $\mathbf{v}$ is the flow velocity, p is the plasma pressure, $\mathbf{B}$ is the magnetic field, $\mathbf{B}$ is the magnetic field, $\mathbf{J}$ is the current density, $\mathbf{g}$ is the gravity force, $\Phi \equiv \mu p \nabla^2 \mathbf{v}$ is the viscosity, $\gamma = 5/3$ is the ratio of specific heats, $\eta = \eta_0 (T/T_0)^{-3/2}$ is the resistivity, $T = p/\rho$ is the temperature, T_0 is the ionospheric temperature, and $\mathbf{B}_d$ is the dipole magnetic field. Typical numerical values are $\eta_0 = 0.01$ and μ/ρ_{sw}. Hence the typical magnetic Reynolds number is $S = \tau_\eta/\tau_A = 100 - 2000$. Here, the subscript sw indicates the quantities in the solar wind. The MHD quantities are decomposed in the z direction as in VPP Fortran. The MHD code, which is fully vectorized and fully parallelized in VPP Fortran, was translated from VPP Fortran to HPF/JA. We summarize below some key points involved in that translation.

To begin with, there is no concept of the global variables in VPP Fortran and also "equivalence" sentence to connect the global variables with local variables cannot be used in HPF/JA. To vectorize for inner do loops and to parallelize for outer do loops is the same in HPF/JA as in VPP Fortran. The efficiency of parallelization increases when the outer do loops are combined in wider range. A part of the 3-dimensional MHD code translated from VPP Fortran to HPF/JA is shown as an example in this section and the MHD code with the complete boundary condition in HPF/JA is seen on a Web page whose address will be shown later in this text. A domain decomposition in the z direction is used and overlap data distributed among PEs are handled by "shadow" and "reflect" in HPF/JA.

Example of our 3-dimensional MHD code translated from VPP Fortran to HPF/JA

```
      parameter (npe=16)
!hpf$ processors pe(npe)
      parameter (nx=500,ny=200,nz=400,nxp=150)
      parameter(n1=nx+2,n2=n1*(ny+2),n3=n2*(nz+2))
      parameter(nb=8,nbb=11,n4=n3*nb,n5=n3*nbb,
     1          n6=n3*18)
c
      dimension f(nx2,ny2,nz2,nb),u(nx2,ny2,nz2,nb),
```

```
     1            ff(nx2,ny2,nz2,nb),p(nx2,ny2,nz2,nbb),
     2            pp(nx2,ny2,nz2,3),fdd(mfd,nfd)
c
!hpf$  distribute f(*,*,block,*) onto pe
!hpf$  distribute u(*,*,block,*) onto pe
!hpf$  distribute pp(*,*,block,*) onto pe
!hpf$  distribute ff(*,*,block,*) onto pe
!hpf$  distribute p(*,*,block,*) onto pe
!hpf$  shadow     f(0,0,1:1,0)
!hpf$  shadow     u(0,0,1:1,0)
!hpf$  shadow     pp(0,0,1:1,0)
!hpf$  shadow     ff(0,0,1:1,0)
!hpf$  shadow     p(0,0,1:1,0)
!hpf$  asyncid id1
c
       do 410 ii=1,itap
       do 174 m=1,nb
       do 174 k=1,nz2
       read(10) f(1:nx2,1:ny2,k,m)
  174  continue
  410  continue
c
       do 100 ii=1,last
c      boundary condition at nz=1 and nz=nz2
       xx4=0.5*hx*float(2*nxp-nx1-2)
       xx3=hhx(nx1)+xx4
c
!hpf$  independent,new(i,j,k,m)
       do 31 k=1,nz2
!hpf$  on home(f(:,:,k,:)),local(f,i,j,m,x,xx,xx3,xx4,
!hpf$*         hhx) begin
       do 30 m=1,nb
       if(k.eq.1) then
       do 311 j=2,ny1
       f(2,j,1,m)=f(1,j,2,m)
       do 311 i=3,nx1
       x=hhx(i)+xx4
       x=1.0-x/xx3
       x=amin1(x,1.0)
       x=1.0+x/3.0
       xx=1.0-x
       f(i,j,1,m)=x*f(i-1,j,2,m)+xx*f(i-2,j,3,m)
  311  continue
       else if(k.eq.nz2) then
       do 312 j=2,ny1
```

```
      f(2,j,nz2,m)=f(1,j,nz1,m)
      do 312 i=3,nx1
      x=hhx(i)+xx4
      x=amin1(x,1.0)
      x=1.0+x/3.0
      xx=1.0-x
      f(i,j,nz2,m)=x*f(i-1,j,nz1,m)+xx*f(i-2,j,nz,m)
  312 continue
      end if
   30 continue
!hpf$ end on
   31 continue
c
!hpf$ asynchronous(id1),nobuffer begin
      f(2:nx1,1,nz2:1:-1,1:nb) = f(2:nx1,2,1:nz2,1:nb)
!hpf$ end asynchronous
!hpf$ asyncwait(id1)
c
!hpf$ reflect f
c
c     first step
c     step of k=k
!hpf$ independent,new(i,j,k)
      do 90 k=1,nz1
!hpf$ on home(u(:,:,k,:)),local(f,u,p,pp,ff,i,j,m,
!hpf$*        x,y,z,hx,hy,hz,xx4,hhx,ar2,
!hpf$*        ro2,ro3,dx1,dy1,dz1,gam,eud,gra,ar1,
!hpf$*        uxd,x1,x2,x3,x5,y1,y5,ro02,pr02,vmax)
!hpf$*        begin
c
c     current
      do 38 j=1,ny1
      do 38 i=1,nx1
      p(i,j,k,11)=0.25*((f(i+1,j+1,k+1,7)+f(i+1,j,k+1,7)
     1  +f(i+1,j+1,k,7)+f(i+1,j,k,7)-f(i,j+1,k+1,7)
     2  -f(i,j,k+1,7)-f(i,j+1,k,7)-f(i,j,k,7))/hx
     3  -(f(i+1,j+1,k+1,6)-f(i+1,j,k+1,6)+f(i+1,j+1,k,6)-
     4      f(i+1,j,k,6)+f(i,j+1,k+1,6)-f(i,j,k+1,6)
     5      +f(i,j+1,k,6)-f(i,j,k,6))/hy)
c
!hpf$ end on
   90 continue
c
      do 173 m=1,nb
      do 173 k=1,nz2
```

```
       write(ntap) f(1:nx2,1:ny2,k,m)
  173  continue
c
```

In the MHD code, all the necessary variables, which appear in the do loop for parallelization, are indicated in the on home(u(:,:,k,:)),local( ) sentence in order to prevent unnecessary communication between PEs (Processing Elements) and to vectorize in the inner do loops. Otherwise, the inner do loops cannot be vectorized. If the inner do loops are not vectorized, we decompose the do loop for parallelization into small do loops and try again to find missing variables in the on home( ), local( ) sentence. It is not difficult to find all the necessary variables in small do loops for parallelization. After the full vectorization is confirmed in each of do loops for parallelization, we combine them to form a wider single do loop for parallelization to obtain higher efficiency. Thus we can accomplish a complete translation of the MHD code from VPP Fortran to HPF/JA

In VPP Fortran, global and local variables are defined and the two variables are linked by "equivalence" sentence as follows,

```
!xocl index partition ind=(pe,index=1:nz2,part=band)
!xocl index partition indo=(pe,index=1:nz2,part=band,
                                              overlap=(1,1))
      dimension f(nx2,ny2,nz2,nb),u(nx2,ny2,nz2,nb),
     1           ff(nx2,ny2,nz2,nb),p(nx2,ny2,nz2,nbb),
     2           pp(nx2,ny2,nz2,3)
      dimension gf(nx2,ny2,nz2,nb),gu(nx2,ny2,nz2,nb),
     1              gff(nx2,ny2,nz2,nb),gpp(nx2,ny2,nz2,3)
!xocl local f(:,:,/indo,:),u(:,:,/indo,:),pp(:,:,/indo,:)
!xocl local ff(:,:,/indo,:),p(:,:,/indo,:)
!xocl global gf,gu,gff,gpp
      equivalence (gf,f),(gff,ff),(gu,u),(gpp,pp)
      common /blk/gf,gpp
c
!xocl overlapfix(f,u) (id)
!xocl movewait (id)
c
c     first step
c     step of k=k
!xocl spread do /ind
      do 90 k=1,nz1
c
   90 continue
!xocl end spread
```

The overlap data at the boundary of distributed data are handled by "sleeve" function in VPP Fortran, which are composed of "index partition" and "overlapfix" sentences. The local variables are used for all the computation and the corresponding global variables are used only in read and write sentences. How-

ever, there is no concept of global variables in HPF and "equivalence" sentence cannot be used. Therefore, we need to prepare a new array for input/output, and lump transmission of distributed data is done for the new array by "asynchronous" sentence. Then we can use the new array for read and write sentences without format, and in doing so retaining a high speed. Of course, the program size increases for the new array. It is noted that lump transmission by "asynchronous" sentence must be used to transfer the distributed data; otherwise it often takes an extremely long time for data transfer. "Asynchronous" sentence has another function to allow lump data transfer of the distributed data with different directions of decomposition. In an example of the final version translated from VPP Fortran to HPF/JA, the unformatted read and write sentences are used for a part of large array of distributed data without arranging a new array.

4 Comparison of Processing Capability Using VPP Fortran and HPF/JA

Our aim was to translate the 3-dimensional global MHD code simulating the interaction between the solar wind and the earth's magnetosphere from VPP Fortran to HPF/JA and to achieve an efficiency more than 50[5]-[7]. In the global MHD simulation, the MHD and Maxwell's equations are solved by the modified leap-frog method as an initial value and boundary value problem to study the response of the magnetosphere to variations of the solar wind and the interplanetary magnetic field (IMF) [4],[5]. Because the external boundary is put in a distant region to minimize its influence on the boundary condition and the tail boundary is extended to permit examination of the structure of the distant tail, higher spatial resolution is required to obtain numerically accurate results. Therefore, the number of 3-dimensional grid points was increased up to the maximum limit of the computer system. Examples of MHD simulations of the solar wind-magnetosphere interaction are shown in Figs. 1 and 2 [5],[7]. Figure 1 shows the 3-dimensional structure of magnetic field lines under steady state conditions for the earth's magnetosphere when the IMF is northward and duskward. A dawn -dusk asymmetry appears in the structure of magnetic field lines because magnetic reconnection at the magnetopause occurs in the high latitude tail on the dusk side in the northern hemisphere and on the dawn side in the southern hemisphere. Figure 2 shows a snapshot of the earth's magnetosphere obtained by a 3-dimensional global MHD simulation at a time when the ACE satellite is monitoring the upstream solar wind and IMF every 1 minute. These data were used as input to the simulation. This simulation represents one of the fundamental studies in the international space weatherindexspace weather program designed to develop numerical models and to understand the variations in the solar-terrestrial environment from moment to moment. It requires 50-300 hours of computation time to carry out these 3-dimensional global MHD simulations of the earth's magnetosphere for 25-150 hours of real time data even using 16PEs of the Fujitsu VPP5000/56.

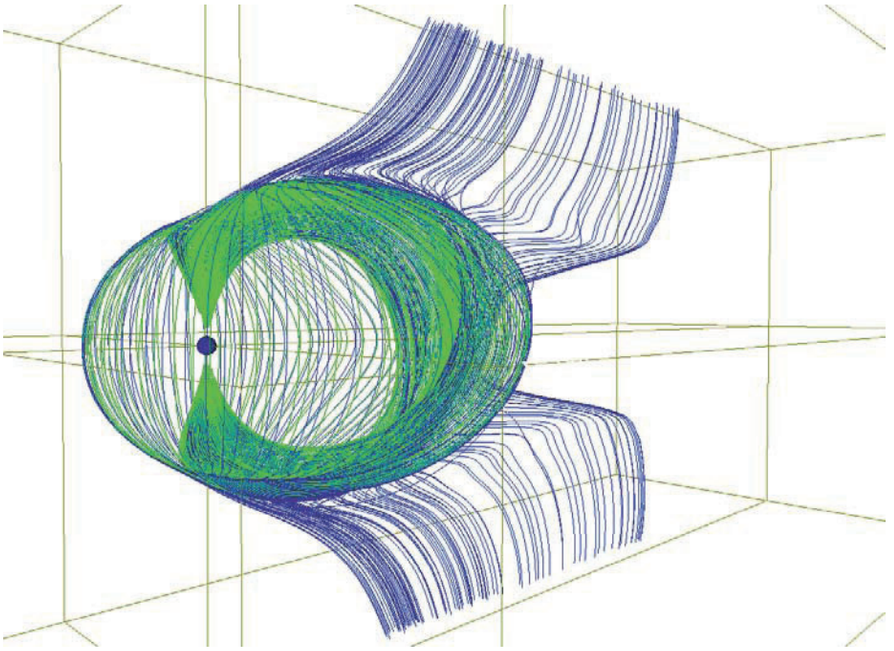


Fig. 1. Structure of 3-dimensional magnetic field lines for steady state conditions in the earth's magnetosphere and an interplanetary magnetic field which is northward and duskward. An asymmetry appears in the magnetosphere due to the occurrence of magnetic reconnection on the dusk side in the northern hemisphere and on the dawn side in the southern hemisphere near the magnetopause.

Table 1 shows the comparison of the processing capabilities of VPP Fortran and HPF/JA for the 3-dimensional global MHD code describing the solar wind-magnetosphere interaction run on the Fujitsu VPP5000/56. Both the Fortran codes are fully vectorized and fully parallelized, and are adjusted to reach the maximum performance using the results of several test runs applying practical simulations. The table includes the number of PEs, 3-dimensional grid points (except for the boundary), CPU time (sec) to execute and advance of one time step, computation speed (Gflops) and computation speed per PE (Gflops/PE) for both VPP Fortran and HPF/JA.

The characteristics of the Fujitsu VPP5000/56 in the Computer Center of Nagoya University are as follows,

Number of PEs: 56 PE
Theoretical maximum speed of PE: 9.6 Gflops
Maximum memory capacity per PE for users: 7.5 GB

As is known from Amdahl's law, it is quite important to fully vectorize for all inner do loops and to fully parallelize for all outer do loops in order to in-

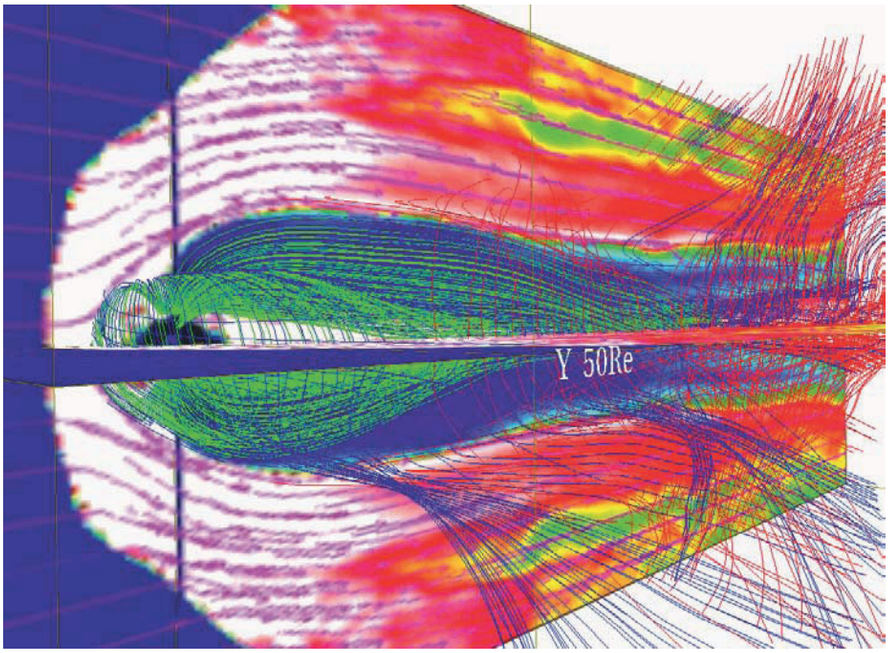


Fig. 2. Examples of snapshots of the earth's magnetosphere obtained by 3-dimensional global MHD simulation, when the observational data of the solar wind and interplanetary magnetic field obtained by the ACE satellite were used as input for the simulation.

crease the efficiency of the vectorization and parallelization performance. The efficiency would be quite low if even only one do loop could not be vectorized or parallelized. Users can achieve excellent computational performance if they can succeed in vectorizing or parallelizing the last do loop. In the 3-dimensional MHD simulation, the number of grid points is given by nx*ny*nz, decomposition of distributed data is taken in z-direction and the size of the array in z-direction becomes nz2=nz+2. Therefore, it is desirable that the array size in the direction of decomposition, nz2 is chosen to be an integer times the number of PEs. Moreover, the efficiency of vectorization and parallelization becomes better as an number of jobs (viz. the number of grid points nx*ny*nz) increases. The computational results for the scalar mode and for the vector mode with only 1PE are also shown in the Table for the sake of comparison with parallel computation. The meaning of this Table can be easily understood if we look at it from that point of view.

It should be noted from the Table that the vector mode with 1PE is about 40 times faster than the scalar mode and that computational speed of HPF/JA is almost comparable to that of VPP Fortran. Furthermore, the scalability of parallel computation is well satisfied given that the computation speed is roughly proportional to the number of PEs. However, when we look at the Table in more detail, it is noted that it was not suitable to choose the grid number to be

Table 1. Comparison of processing capabilities of VPP Fortran and HPF/JA in a 3-dimensional global MHD code for the solar wind-magnetosphere interaction using a Fujitsu VPP5000/56.

Number of PE	Number of grids	VPP Fortran cpu time Gflops Gf/PE	HPF/JA cpu time Gflops Gf/PE
1PE	200x100x478	119.607 (0.17) 0.17	(scalar)
1PE	200x100x478	2.967 (6.88) 6.88	3.002 (6.80) 6.80
2PE	200x100x478	1.458 (14.01) 7.00	1.535 (13.30) 6.65
4PE	200x100x478	0.721 (28.32) 7.08	0.761 (26.85) 6.71
8PE	200x100x478	0.365 (55.89) 6.99	0.386 (52.92) 6.62
16PE	200x100x478	0.205 (99.38) 6.21	0.219 (93.39) 5.84
32PE	200x100x478	0.107 (191.23) 5.98	0.110 (186.13) 5.82
48PE	200x100x478	0.069 (297.96) 6.21	0.074 (276.96) 5.77
56PE	200x100x478	0.064 (319.53) 5.71	0.068 (299.27) 5.34
1PE	500x100x200	2.691 (7.94) 7.94	2.691 (7.94) 7.94
2PE	500x100x200	1.381 (15.47) 7.73	1.390 (15.37) 7.68
4PE	500x100x200	0.715 (29.97) 7.47	0.712 (29.99) 7.50
8PE	500x100x200	0.398 (53.65) 6.71	0.393 (54.38) 6.80
16PE	500x100x200	0.210 (101.87) 6.37	0.202 (105.74) 6.61
32PE	500x100x200	0.131 (163.55) 5.11	0.120 (175.50) 5.48
48PE	500x100x200	0.100 (214.48) 4.46	0.091 (231.69) 4.82
56PE	500x100x200	0.089 (239.48) 4.28	0.086 (244.85) 4.37
4PE	800x200x670	7.618 (30.06) 7.52	8.001 (28.62) 7.16
8PE	800x200x670	3.794 (60.36) 7.54	3.962 (57.81) 7.23
12PE	800x200x670	2.806 (81.61) 6.80	3.005 (76.21) 6.35
16PE	800x200x670	1.924 (119.00) 7.44	2.012 (113.85) 7.12
24PE	800x200x670	1.308 (175.10) 7.30	1.360 (168.44) 7.02
32PE	800x200x670	0.979 (233.85) 7.31	1.032 (221.88) 6.93
48PE	800x200x670	0.682 (335.62) 6.99	0.721 (317.80) 6.62
56PE	800x200x670	0.595 (384.61) 6.87	0.628 (364.87) 6.52
16PE	1000x500x1118	9.668 (123.52) 7.72	9.619 (125.50) 7.84
32PE	1000x500x1118	5.044 (236.73) 7.40	4.992 (241.83) 7.56
48PE	1000x500x1118	3.550 (336.40) 7.01	3.479 (346.97) 7.23
56PE	1000x500x1118	2.985 (400.04) 7.14	2.935 (411.36) 7.35
32PE	1000x1000x1118	9.979 (239.33) 7.48	9.813 (243.37) 7.61
48PE	1000x1000x1118	7.177 (332.79) 6.93	7.028 (339.85) 7.08
56PE	1000x1000x1118	5.817 (410.55) 7.33	5.794 (412.23) 7.36

: Gflops is an estimated value for comparison with a computation by 1 processor of a CRAY Y-MP C90

nx*ny*nz=500*100*200 as used in the previous simulation for parallel computation. This is because the ratio of the number of array variables in the direction of decomposition (nz2=202=101*2) to the number of PEs is not normally an integer. This follows from the fact that the computation speed increases, but is lower than the speed expected from linear proportionality to the number of PEs.

The scalability is greatly improved when we choose the number of grid points as nx*ny*nz=200*100*478 (nz2=480=2**5*3*5) and 800*200*670 (nz2=672=2**5*3*7) in the MHD simulation. Furthermore we achieved a rather high performance exceeding 400 Gflops in both cases of VPP Fortran and HPF/JA when we chose a large number of grid points given by nx*ny*nz=1000*500*1118 and 1000*1000*1118. This is because nz2 is an integer times the number of PEs and also the program size of MHD code became larger. These results clearly show that a 3-dimensional global MHD simulation of the earth's magnetosphere can be performed at a speed of over 400 Gflops with an efficiency of 76.5vector and parallel computation that permitted comparison with catalog values.

Concrete examples of the MHD codes translated from VPP Fortran to HPF/JA are not shown in the present paper, however a part of the boundary condition in the HPF/JA MHD code and a test program of the 3-dimensional wave equation can been seen on a Web page which will be shown later.

Some important and general points in using the HPF/JA in Fujitsu VPP5000 are as follows:

The first point refers to the local sentence. If all the parameters are removed in local sentence of "on home local( )", it is equivalent to arranging all the variables and array from "begin" to "end on" to be completely written in the local sentence. Therefore it is convenient to use the function in the "do loop" sentence for the fully vectorized and fully parallelized "do loop" in the VPP Fortran program. The second point refers to read and write sentences. If distributed array data are not very numerous, it is convenient to use unformatted read and write sentences such as

```
      write(*,100) a,b,c
      write(10) a,b,c
```

which is simple description and for which the execution time is not long. Moreover, we can also use the following unformatted read and write sentences for a part of large array of distributed data without arranging a new array.

```
      do 174 m=1,nb
      do 174 k=1,nz2
      read(10) f(1:nx2,1:ny2,k,m)
 174  continue

      do 176 m=1,nb
      do 176 k=1,nz2
      write(ntap) f(1:nx2,1:ny2,k,m)
 176  continue
```

This method does not require any increase of the program size in input and output for a large array of distributed data and the execution time is almost comparable to that a newly arranged array is prepared to use a copy function by asynchronous sentence. Therefore, we can keep the program size of HPF/JA, where "equivalence" sentence cannot be used to link global and local variables, the same as for VPP Fortran.

5 Development of MHD Simulation

Since 2 and 3-dimensional global MHD simulations of the interaction between the solar wind and the earth's magnetosphere have been carried out over the past 20 years [1]-[6],[13],[14], it is useful to present a short history of their development and evaluate their future. Figure 3 shows a summary plot of the 2 and 3-dimensional global MHD simulations indicating how many grid points were used as the years went by. In this figure, the black circles and triangles show the MHD simulations carried out by other groups, while the white ones and crosses show the MHD simulations carried out by our group. We began the 3-dimensional MHD simulations to study MHD instabilities in toroidal plasmas using small number of grid points (given by 18*18*12 and 26*26*18 including

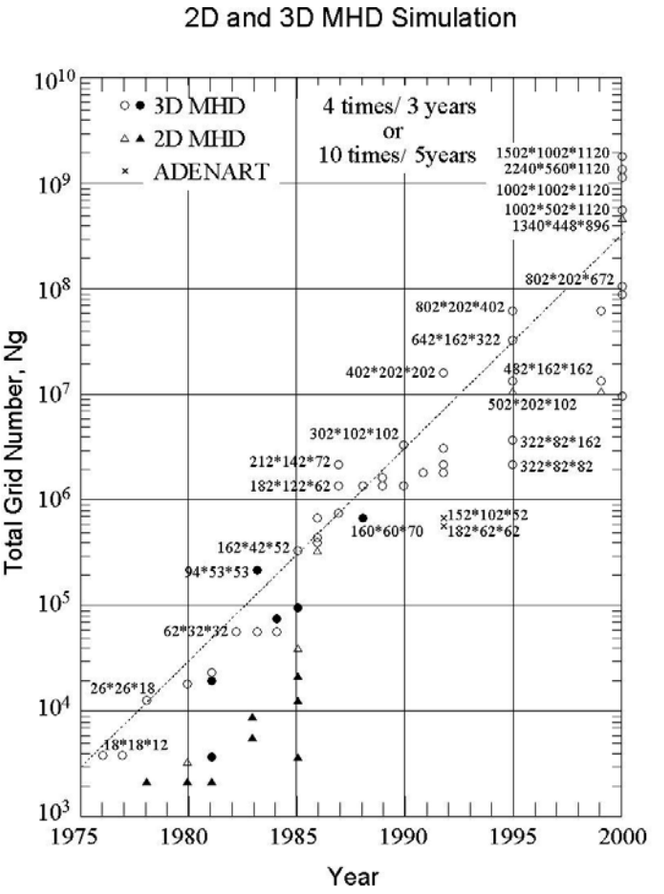


Fig. 3. Development of 2 and 3-dimensional MHD simulations as quantified by the total number of grid points used; the total number of grid points has increased 4 fold every 3 years or 10 fold every 5 years.

the boundary) in 1976 [13],[14]. However, the MHD simulations required about 50-200 hours of computing time using Fujitsu M-100 and M-200 computers. The execution of such large scale MHD simulations was practically limited by speed of the computers at that time.

We began a global MHD simulation of the earth's magnetosphere using a larger number of grid points (given by 62*32*32 and 50*50*26, including the boundary) using a CRAY-1 in 1982 [1]-[3], when the memory of CRAY-1 of 1MW limited the scale of the MHD simulation. At that time, we considered how many interesting simulations could be executed if 100 cube grid points could be used. The Japanese supercomputers subsequently appeared, and we could carry out the MHD simulations with 100 cube grid points using Fujitsu VP-100, VP-200 and VP-2600 machines. Later, we had anopportunity to use the Fujitsu VPP500 vector parallel machine in 1995 and began to use the VPP5000 machine in 2000; this practically allowed us to carry out the MHD simulation with 1000 cube grid points. When we look at the development of the MHD simulations from the point of view of the number of grid points used over those 24 years, we can recognize the pattern of a 4 fold increase every 3 years or a 10 fold increase every 5 years.

Figure 4 shows the computation time for an advance of one time step which is needed to execute the MHD simulation, while Fig. 5 shows the capacity of the

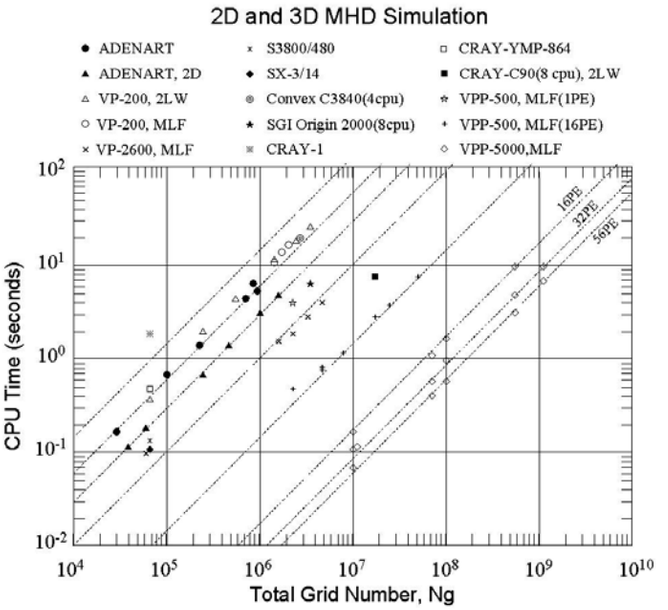


Fig. 4. Comparison of computation times corresponding to an advance of one time step versus total number of grid points in the MHD simulation codes.

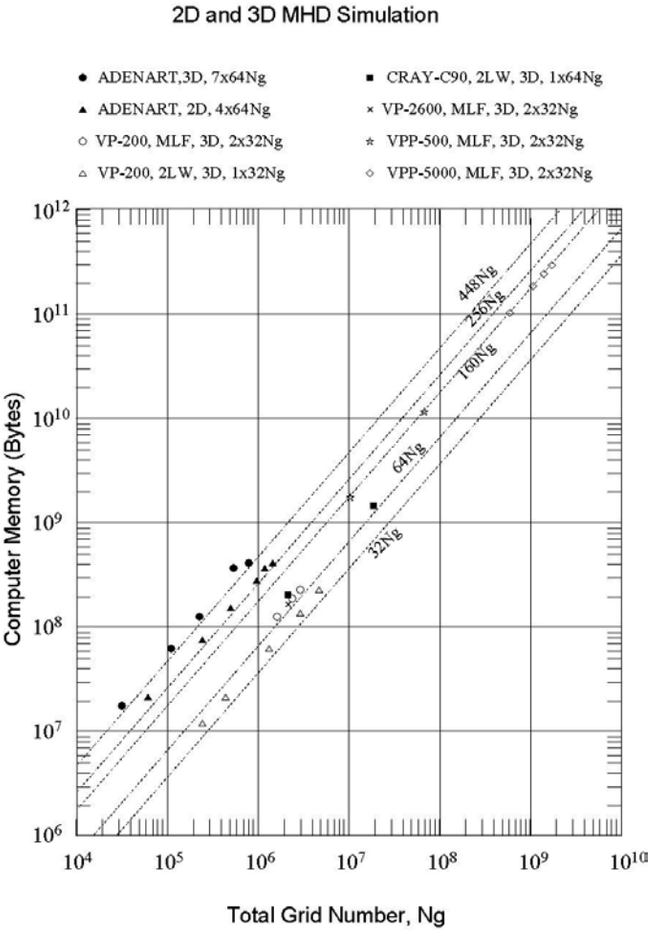


Fig. 5. Comparison of total capacity of the necessary memory versus total number of grid points in the MHD simulation codes.

computer memory which is needed to execute the MHD simulation along with the total number of grid points.

The computation time and capacity generally depend of the kind of computer and numerical methods, and the computation time becomes longer and the required computer memory increases in linear proportion to the number of total grid points. Since the global MHD simulation of the earth's magnetosphere requires repetitive calculations from several thousand to tens of thousands times the one time step advance, it is necessary to make the computation time of one time step advance less than 10 seconds in order to carry out practical MHD simulations [5],[6],[8]. Thus, the number of usable grid points is automatically limited in MHD simulations by the speed of the computer. At the same time,

the program size for MHD simulations with the chosen number of grid points must be less than size of the computer memory.

Since the time evolution of 8 physical components (i.e. the density, 3 components of velocity, plasma pressure and 3 components of magnetic field) is calculated in the MHD code, the number of independent variables is 8 Ng and the capacity is 32 Ng Bytes in single precision variables, where Ng is the total number of grid points. Therefore, the key variables in the MHD code are 32 Ng Bytes using the 2 step Lax-Wendroff method (2LW) and 64 Ng Bytes using the modified Leap-Frog method (MLF) because the MHD quantities must be stored for two consecutive time steps. The 3-dimensional global MHD code which was originally developed for the CRAY-1 and has been used by other vector machines requires only 1.3 times the capacity to store the MHD key variables for a 1-dimensional array in order to effectively decrease the work area. It becomes impossible to effectively decrease the work area for parallel computers, as a result the program size increases considerably in parallel machines such as ADENART, VPP500 and VPP5000.

However, the required memory in VPP500 and VPP5000 is 2.5 times the available memory of the computer to store the key variables (160 Ng $= 2.5 * 64$ Ng) in MLF. If the dependence on required memory for the total grid number could be concretely demonstrated, we could easily estimate how large an MHD simulation could be executed to study specified subjects and could establish more precisely the prospects for the future.

Figure 6 shows a comparison of the computer processing capability with the total grid number in MHD simulations. Rapid progress has been made in increasing performance by moving from vector machines to vector parallel machines. Newer high performance workstations and personal computers have the capability of older supercomputers (speed of about 0.3 Gflops). However, the processing capability of high end supercomputers is still about a thousand times higher than that of other computers at all times. Moreover, definitive differences between supercomputers and other computers appear in the shortage of main memory and cache required to carry out MHD simulations with a large number of grid points. It is practically impossible to execute a 3-dimensional MHD simulation code using workstations and personal computers when the fully vectorized and fully parallelized MHD code would need CPU times of 100 hours using a supercomputer such as the Fujitsu VPP5000. It is nowadays necessary to use a supercomputer with a maximum performance capability in studies involving 3-dimensional global MHD simulations of the earth's magnetosphere.

6 Conclusions

We have carried out a 3-dimensional global magnetohydrodynamic (MHD) simulation of the interaction between the solar wind and the earth's magnetosphere on a variety of computers using a fully vectorized MHD code. However, simulation scientists have lost their common language with the appearance of vector parallel and massive parallel supercomputers. We sincerely hope to see the development

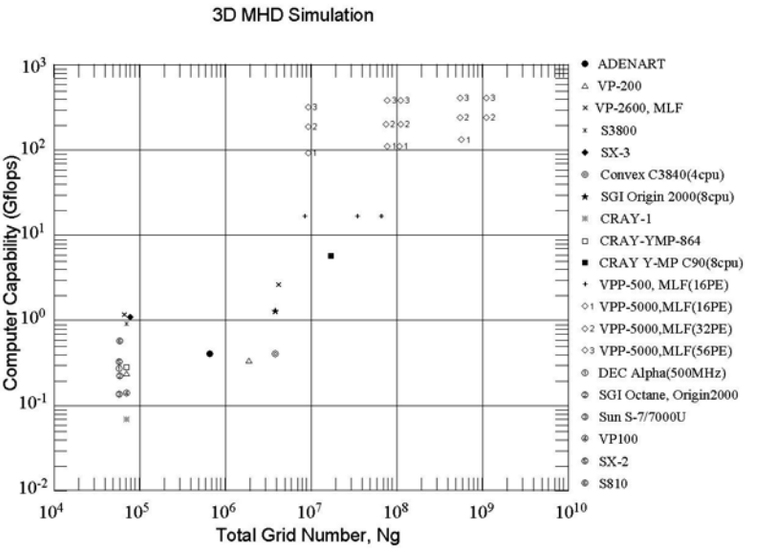


Fig. 6. Comparison of the processing capability versus total number of grid points in the MHD simulation codes.

of a common language in the supercomputer world. The candidates appear to be High Performance Fortran (HPF) and Message Passing Interface (MPI). We look forward to the time when we can use these compilers in supercomputers. We have had the opportunity to use HPF/JA with the Fujitsu VPP5000/56 supercomputer in the Computer Center of Nagoya University since June 2000. We translated our 3-dimensional MHD simulation code for the solar-terrestrial interaction from VPP Fortran to HPF/JA and the code was fully vectorized and fully parallelized in VPP Fortran. The performance of the HPF MHD code was almost comparable to that of the VPP Fortran in a typical simulation using a large number (56) of Processing Elements (PEs). We have reached the conclusion that fluid and MHD codes that are fully vectorized and fully parallelized in VPP Fortran can be relatively easily translated to HPF/JA, and a code in HPF/JA can be expected to achieve comparable performance to one written in VPP Fortran.

The 3-dimensional global MHD simulation code for the earth's magnetosphere solves the MHD and Maxwell's equations in 3-dimensional Cartesian coordinates (x, y, z) as an initial and boundary value problem using a modified leap-frog method. The MHD quantities are distributed in the z direction. The quantities in the neighboring grids can be calculated by the HPF/JA instruction sentences "shadow" and "reflect" when they exist in another PE. Unnecessary communication among PEs can be completely avoided by the instructions "independent, new," and "on home, local." Moreover, lump transmission of data is used in the calculation of the boundary conditions by the instruction "asynchronous." It was not necessary to change the fundamental structure of the MHD

code in the translation procedure. This was a big advantage in translating the MHD code from VPP Fortran to HPF/JA.

Based on this experience, it is anticipated that we will find little difficulty in translating programs from VPP Fortran to HPF/JA, and we can expect an almost comparable performance on the Fujitsu VPP5000. The maximum speed of the 3-dimensional MHD code was over 230 Gflops for 32 PEs and over 400 Gflops for 56 PEs. We hope that the MHD code rewritten in HPF/JA can be executed on other supercomputers such as the Hitachi SR8000 and the NEC SX-5 in near future and that HPF/JA becomes a useful common language in the supercomputer world. We believe that this is a necessary condition to restart a new and higher level of collaborative research on 3-dimensional global MHD simulation of the magnetosphere with other scientists around the world and expect to see efforts by simulation scientists as well as by the Japanese supercomputer companies to reach this goal.

We have made available a part of the boundary condition in the HPF/JA MHD code and a test program of the 3-dimensional wave equation at the following Web address:

http://gedas.stelab.nagoya-u.ac.jp/ simulation/hpfja/hpf01.html.

Acknowledgments

We would like to thank the staff of the Computer Center of Nagoya University and the SEs of the Fujitsu company for their helpful discussions and support. We thank G. Rostoker for useful comments. This work was supported by a grant in aid for science research from the Ministry of Education, Science and Culture. Computing support was provided by the Computer Center of Nagoya University.

References

1. Ogino T., R.J. Walker, M. Abdalla and J.M. Dawson, An MHD simulation of By-dependent magnetospheric convection and field-aligned currents during northward IMF, J. Geophys. Res. 90, 10835-10842, 1985.
2. Ogino T., A three-dimensional MHD simulation of the interaction of the solar wind with the earth's magnetosphere: The generation of field-aligned currents, J. Geophys. Res., 91, 6791-6806, 1986.
3. Ogino T., R.J. Walker, M. Abdalla and J.M. Dawson, An MHD simulation of the effects of the interplanetary magnetic field By component on the interaction of the solar wind with the earth's magnetosphere during southward interplanetary magnetic field, J. Geophys. Res., 91, 10029-10045, 1986.
4. Ogino T., R.J. Walker and M. Ashour-Abdalla, A global magnetohydrodynamic simulation of the magnetosheath and magnetopause when the interplanetary magnetic field is northward, IEEE Transactions on Plasma Science, Vol.20, No.6, 817-828, 1992.
5. Ogino T., R.J. Walker and M. Ashour-Abdalla, A global magnetohydrodynamic simulation of the response of the magnetosphere to a northward turning of the interplanetary magnetic field, J. Geophys. Res., Vol.99, No.A6,11,027-11,042, 1994.

6. Ogino, T., R.J. Walker, and M.G. Kilvelson, A global magnetohydrodynamic simulation of the Jovian magnetosphere, J. Geophys. Res., 103, 225-235, 1998.
7. Ogino, T., Magnetohydrodynamic simulation of interaction between the solar wind and magnetosphere, J. Plasma and Fusion Research, (Review paper), Vol.75, No.5, CD-ROM 20-30, 1999.
8. Ogino, T., Computer simulation of solar wind-magnetosphere interaction, Center News, Computer Center of Nagoya University, Vol.28, No.4, 280-291, 1997.
9. Fujitsu Co., VPP Fortran Programming, UXP/V, UXP/M Ver.1.2, April, 1997.
10. Tsuda, T., On usage of supercomputer VPP5000/56, Center News, Computer Center of Nagoya University, Vol.31, No.1, 18-33, 2000.
11. High Performance Fortran Forum, Official Manual of High Performance Fortran 2.0, Springer, 1999.
12. Fujitsu Co., HPF programming -parallel programming (HPF version)-, UXP/V Ver.1.0, June, 2000.
13. Ogino T., H. Sanuki, T. Kamimura and S. Takeda, Nonlinear evolution of the resistive MHD modes in a toroidal plasma, J. Phys. Soc. Japan, 50, 315-323, 1981.
14. Ogino T., H. Sanuki, T. Kamimura and S. Takeda, Nonlinear evolution of the kink ballooning mode in a toroidal plasma, J. Phys. Soc. Japan, 50, 1698-1705, 1981.

Numerical Schemes for the Analysis of Turbulence – A Tutorial

Thierry Dudok de Wit

LCPE, UMR 6115 CNRS-Université d'Orléans, 3A, av. de la Recherche Scientifique, F-45071 Orléans cedex 2, France

Abstract. The analysis of plasma turbulence has traditionally relied on limited repertoire of methods such as Fourier analysis, correlation analysis, etc. This text gives a short overview of what deeper insight can be obtained by using techniques that exploit nonlinear properties of the data. It covers both higher order spectra and higher order statistics.

1 Linear vs Nonlinear World

Although linear descriptions traditionally constitute a rule of conduct in the analysis of space plasma data, there is quite some evidence that nonlinear descriptions may provide considerable added value. Linear descriptions can all be cast within the single convenient framework of Fourier analysis, which includes well known quantities such as power spectra and autocorrelation functions. There is no such single framework, however, for making nonlinear descriptions. A key problem in nonlinear world therefore is the choice of the right technique, or equivalently, the right invariant. Physical guidance and a good understanding of what the model produces are a must.

The repertoire of techniques for analysing nonlinear phenomena is huge and increases by the day, see for example [1,36,61]. Most of the texts on turbulence are essentially devoted to neutral fluids only [49,44,26,9]. As it turns out, only a subset of these techniques is really relevant for space plasma applications. By relevant we mean that a clear physical interpretation can be given to the output, and a compatibility between the underlying assumptions and properties of the model (or the observations).

Several applications to problems encountered in space plasmas can nevertheless be found in [24,28,22,75,76]. Most of them deal with experimental data, but the extension to simulation data is immediate. Significant advantages of the latter are generally a smaller noise level, a more flexible size of the dataset and the possibility to access quantities that cannot always be measured. If simulation data have received comparatively less attention so far, it is not by lack of relevance, but often because we simply don't know yet how to properly handle large ensembles of multivariate data !

There exist essentially three major families of techniques for analysing turbulence, all of which are rooted in the theory of statistics and that of dynamical systems. They are respectively:

J. Büchner, C.T. Dum, M. Scholer (Eds.): LNP 615, pp. 315–343, 2003.

Higher order spectra (Sect. 2), which are a generalization of Fourier analysis to include information about phase coherence. This approach is particularly relevant for wave phenomena that weakly depart from nonlinearity. Typical applications are: nonlinear wave interactions, wave-particle interactions, weak turbulence, etc.

Higher order statistics (Sect. 3), although closely related to higher order spectra, are more generally applied to stochastic processes, regardless of the degree of nonlinearity. Typical applications are fully developed turbulence, fluctuations in complex systems, etc.

Phase space techniques is the name given to techniques that are generally applied to chaotic systems, whose time evolution can be approximated with a low-dimensional deterministic model. Such systems are not generic in space plasmas but the concepts that have been developed in this framework are powerful and deserve attention. Two recent textbooks are [36,9].

Phase space methods will not be considered here because would require a textbook on their own. Our selection is therefore somewhat arbitrary, even more so since the same concept can often be found in more than one family.

The domain of application of each family can be characterized by comparing the degrees of stochasticity (*how much is the dynamics ruled by external noise terms?*) and nonlinearity (*how nonlinear is the process?*), see Fig. 1.

Notice how the applicability of each technique is usually confined to a neighbourhood around some hypothesis. One could of course add more dimensions to the plot by introducing the extension (*is the process temporal or spatio-temporal?*), the number of degrees of freedom, etc.

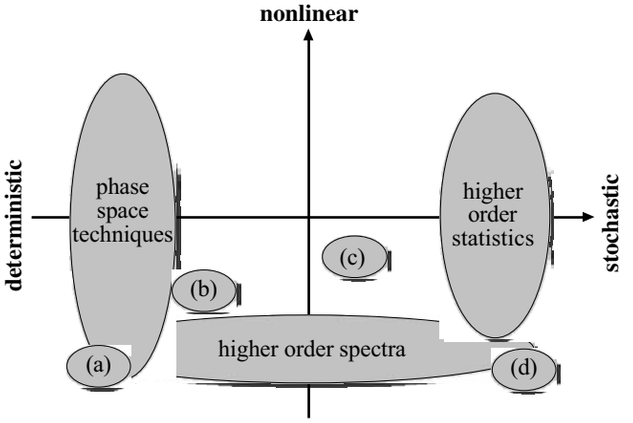


Fig. 1. Sketch of the variety of systems spanned by the properties "nonlinearity" and "stochasticity". Besides the two main families that are described in this text, there exist many other techniques based for example on linear oscillations (a), nonlinear Fourier transforms for solitary waves (b), hidden Markov models (c), and stochastic linear models (d). Sketch made after a figure from [61].

2 Higher-Order Spectra and Spectral Energy Transfers

One of the most obvious approaches for making the transition from a linear to a nonlinear description, is to start from a linear description of the system of interest, and subsequently introduce a weak nonlinearity. How far this perturbative approach can be extended to strongly nonlinear systems depends on the type of problem one is addressing.

Since one starts with a linear description, it is appropriate to describe the system in terms of its eigenmodes, namely Fourier modes. The basic assumption underlying standard Fourier analysis is that any stationary fluctuating physical quantity can be regarded as the superposition of statistically independent Fourier modes. Accordingly, all the relevant information is contained in the amplitude of these modes only, that is, the power spectral density (or alternatively, the autocorrelation function). If, however, there exists some parametrical or nonlinear physical process, then the phases of the Fourier modes are not independent anymore, and information is also conveyed by the phases. Higher order spectra[1] provide a means for characterizing such phase coherence [48,53,40].

2.1 Why Higher Order Spectra?

Let us start from an example with magnetic field data measured by the AMPTE UKS spacecraft just upstream the Earth's quasiparallel bow shock (see [20] for more details). In this dataset, the turbulent wavefield is characterized by the occasional occurrence of large-amplitude structures that are preceded at their upstream edge by whistler wavetrains, see Fig. 2. Two questions arise: 1) is there a causal relationship between the whistler waves and the large structures? 2) If such a coupling exists, are whistlers just instabilities that are triggered by the large structures or do they actually grow out of them? Higher order spectra can answer the first question, and spectral energy transfers the second.

Higher order spectra are usually defined in the field of signal processing, using cumulants [11]. Two applications to laboratory plasma experiments can be found in [37,68]. Let us start with a nonlinear system, whose dynamics is a function of time t and space x, and is described by the generic model

$$\frac{\partial u(x,t)}{\partial x} = f(u(x,t)) \ , \tag{1}$$

where $f(\cdot)$ is a continuous nonlinear and time-independent function, and $u(x,t)$ is the physical quantity of interest[2]. We shall henceforth assume that $u(t)$ is stationary in time and has zero mean. By analogy with Taylor series let us decompose f into a series of linear, quadratic, cubic, and higher order functions

$$f(u) = f_1(u) + f_2(u) + f_3(u) + \cdots \tag{2}$$

[1] higher order spectra are also called polyspectra or multispectra

[2] the spatial derivative is used here for notational convenience; one could switch time and space.

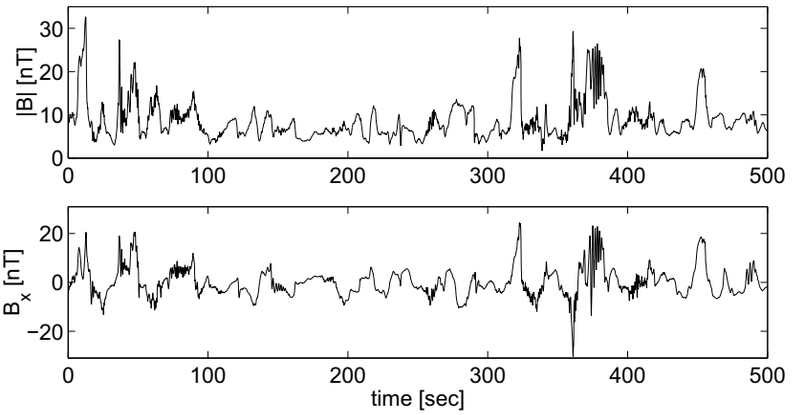


Fig. 2. Excerpt of the AMPTE magnetic field data: modulus of the magnetic field (top) and component along direction of maximum variance (bottom).

.

so that $f_k(\lambda u) = \lambda^k f(u)$. In a weakly nonlinear system, $f_1(u)$ is the term that rules the dynamics. Wiener has shown that with mild assumptions, eq. 2 can be written as a Volterra series.

$$\begin{aligned}\frac{\partial u(x,t)}{\partial x} = &\int g(\tau_1) \ : u(x,t-\tau_1) \ : \mathrm{d}\tau_1 \\ &+ \iint g(\tau_1,\tau_2) \ : u(x,t-\tau_1)\, u(x,t-\tau_2) \ : \mathrm{d}\tau_1 \ : \mathrm{d}\tau_2 \\ &+ \iiint g(\tau_1,\tau_2,\tau_3)\, u(x,t-\tau_1) : u(x,t-\tau_2) : u(x,t-\tau_3)\, \mathrm{d}\tau_1 : \mathrm{d}\tau_2 : \mathrm{d}\tau_3 \\ &+ \cdots\end{aligned} \tag{3}$$

As we shall see later, each term can be ascribed to a different physical process. Let us now take the discrete Fourier transform in time of this series, with the notation $u_p = u(x, \omega_p)$:

$$\frac{\partial u_p}{\partial x} = \Gamma_p u_p + \sum_{k,l} \Gamma_{kl}\, u_k u_l\, \delta_{k+l,p} + \sum_{k,l,m} \Gamma_{klm}\, u_k u_l u_m\, \delta_{k+l+m,p} + \cdots \tag{4}$$

where $\delta_{k,l}$ is the Dirac delta function. Note how the nonlinearity couples each Fourier mode u_k to all other modes. This coupling can only occur in a specific way: for quadratic nonlinearities (terms with the Γ_{kl} kernels), the resonance condition for interacting Fourier modes reads

$$\omega_k + \omega_l = \omega_p \ , \tag{5}$$

whereas for cubic nonlinearities we must have

$$\omega_k + \omega_l + \omega_m = \omega_p \ . \tag{6}$$

Frequencies can be both positive and negative, so for example $\omega_k + \omega_l - \omega_p = 0$ could be rewritten as $\omega_k + \omega_l = \omega_p$

A major asset of this description is the existence of analytical expressions and consequently the possibility to interpret the Volterra kernels Γ in terms of physical processes. Much work has been done in the framework of Hamiltonian systems [78] and several applications to weak plasma turbulence have been reported in [35,50]. With some additional assumptions, even the description of strong plasma turbulence can be considered, see [33].

Volterra series may in principle contain an infinite number of terms, but in many applications (and by definition in all weakly nonlinear systems) only low order terms are significant. The quadratic Volterra kernel Γ_{kl} describes *three-wave interactions* because three different Fourier modes are involved in the resonance condition $\omega_k + \omega_l = \omega_p$. Two examples of such interactions are harmonic generation (the fundamental gives rise to a first harmonic) and the decay instability (a Langmuir wave decays into another Langmuir wave and an ion sound wave). In the same way, the cubic kernel Γ_{klm} describes *four-wave interactions*. The modulational instability is example of a four-wave interaction, in which a Langmuir wave decays into two other Langmuir waves plus a low-frequency ion sound wave.

2.2 Defining Higher Order Spectra

Volterra kernels Γ embody all the information about the nonlinear dynamics of a process and so they should be the prime quantities of interest. One can readily see, however, that their estimation necessitates data with both good sampling in space and in time. Two other quantities called *higher order spectra* and *energy transfer functions* are often preferred, because they can be obtained at lower cost.

Let us first multiply eq. 4 by u_p^* and calculate its expectation[3]. The following series results

$$\langle \frac{\partial u_p}{\partial x} u_p^* \rangle = \Gamma_p \langle |u_p|^2 \rangle + \sum_{k+l=p} \Gamma_{kl} \langle u_k u_l u_{k+l}^* \rangle + \sum_{k+l+m=p} \Gamma_{klm} \langle u_k u_l u_m u_{k+l+m}^* \rangle + \cdots \tag{7}$$

In a homogeneous plasma the left term vanishes, and so the amplitude $\langle |u_p|^2 \rangle$ is not an invariant quantity anymore. This quantity is nothing but the power spectrum

$$P(\omega_p) = \langle u_p u_p^* \rangle \ . \tag{8}$$

The quadratic term is called *bispectrum*

$$B(\omega_k, \omega_l) = \langle u_k u_l u_{k+l}^* \rangle \ , \tag{9}$$

[3] in practice the ensemble averaging can often be replaced by an averaging over time.

and the cubic term *trispectrum*[4]

$$T(\omega_k, \omega_l, \omega_m) = \langle u_k u_l u_m u^*_{k+l+m} \rangle \ . \tag{10}$$

Higher order spectra can therefore be seen as generalizations of the Fourier power spectrum. The bispectrum measures the amount of phase coherence between three Fourier modes that obey the frequency summation rule $\omega_k + \omega_l = \omega_p$. The bispectrum vanishes unless the phases of the modes are correlated, i.e. $B(\omega_k, \omega_l)$ has zero expectation unless the phase difference $\arg u_k + \arg u_l - \arg u_{k+l}$ is constant. Trispectra in the same way measure the amount of phase coherence between four modes.

Figure 3 illustrates this concept of phase coherence for a numerical simulation of a decay instability occurring in a system of three Langmuir waves and one ion-sound wave [42]. The phase of the each wave, taken independently, behaves essentially in a random way, but the phase difference remains almost constant.

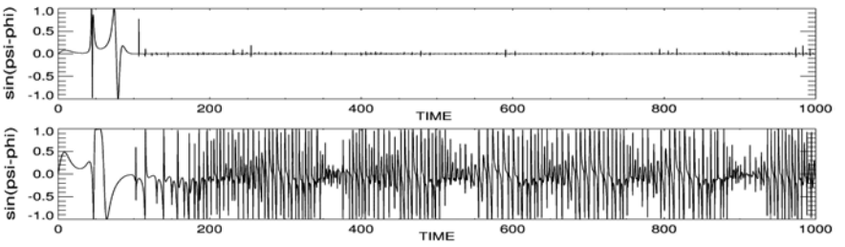


Fig. 3. Sine of the phase difference between three resonant Langmuir waves. After an initial transient, the phase difference settles down to a constant value (upper panel). If the system is slightly detuned from resonance, then the phase difference on average still remains close to zero (lower panel). Figure taken from [42].

It is often more convenient to use normalized quantities. Such a normalization can be performed in different ways (see for example [38]). The usual solution is based on Schwartz's inequality. This leads to a normalized bispectrum, called *bicoherence*

$$b^2(\omega_k, \omega_l) = \frac{|B(\omega_k, \omega_l)|^2}{\langle |u_k u_l|^2 \rangle \langle |u_{k+l}|^2 \rangle} \ . \tag{11}$$

and a normalized trispectrum, called *tricoherence*

$$t^2(\omega_k, \omega_l, \omega_m) = \frac{|T(\omega_k, \omega_l, \omega_m)|^2}{\langle |u_k u_l u_m|^2 \rangle \langle |u_{k+l+m}|^2 \rangle} \ . \tag{12}$$

Both quantities are real and bounded by 0 and 1.

[4] the true definition of the trispectrum is not a fourth order *moment* like here, but a fourth order *cumulant* $T(\omega_k, \omega_l, \omega_m) = \langle u_k u_l u_m u^*_{k+l+m} \rangle - \langle u_k u_l \rangle \langle u_m u^*_{k+l+m} \rangle - \langle u_k u_m \rangle \langle u_l u^*_{k+l+m} \rangle - \langle u_k u^*_{k+l+m} \rangle \langle u_l u_m \rangle$

The interpretation of the bicoherence issues from that of the bispectrum: it measures the proportion of the signal energy at any bifrequency (ω_k, ω_l) that is quadratically phase coupled to ω_{k+l}. The tricoherence similarly quantifies cubic phase couplings between triplets $(\omega_k, \omega_l, \omega_m)$. Such phase couplings are a hallmark of nonlinearity, which is the main motivation for using higher order spectra.

Several applications of bicoherence to space plasmas have been reported, such as Langmuir wave coalescence in the solar wind [6], parametric instabilities in the ionosphere [65], the interaction of radio emitters with the ionosphere [41], nonlinear structures in the magnetosphere [20], and simulations of beam-plasma interactions [66]. Tricoherent analysis so far has only been reported for weak turbulence simulations [38,72] and in magnetospheric turbulence [23].

2.3 Some Properties of Higher Order Spectra

Higher order spectra are intimately connected to higher order statistics. If a time series has a Gaussian probability density, then all higher order spectra are automatically equal to zero. Conversely, nonlinear wave interactions necessarily give rise to non-Gaussian probability densities. The choice of the technique essentially depends on the type of coupling: if few Fourier modes are coupled, then higher order spectra are appropriate since the deviation from Gaussianity may be weak. Conversely, if the coupling involves many different modes (e.g. in fully developed turbulence) then higher order spectra will be small whereas the probability may significantly depart from a Gaussian.

Higher order spectra have been defined so far by considering a single quantity $u(x,t)$, but they may be extended to study the phase coupling between different variables. To distinguish the two situations one should use the prefix *auto-* for a single quantity (i.e. the autobicoherence) and *cross-* for multiple quantities. Beam-plasma interactions is a typical example in which the cross-bicoherence is appropriate. According to the Zakharov equations the formation of cavitons occurs via a coupling between the electron density n and the electric field E, with $n \sim E^2$. To characterize this coupling, one must use the cross-bicoherence $B_{nEE}(\omega_k, \omega_l) = \langle E_k E_l n^*_{k+l} \rangle$ and not the autobicoherence, see for example [72].

The representation of higher order spectra can be a problem, since these quantities depend on several variables. Fortunately, one can take advantage of symmetry properties and strongly reduce the non-redundant frequency domain (called principal domain). The principal domain of the bicoherence is shown in Fig. 4 for a real signal. For the tricoherence it is a prism in 3D [52].

Let us now illustrate these concepts with two examples:

Example 1: We reconsider the AMPTE magnetic field data from Sect. 2.1. The turbulent magnetic field is known to contain nonlinear structures that steepen and progressively decay into dispersive whistler wave packets. This decay occurs via nonlinear wave interactions, and thus gives rise to phase couplings that should appear in the wavefield autobicoherence and autotricoherence.

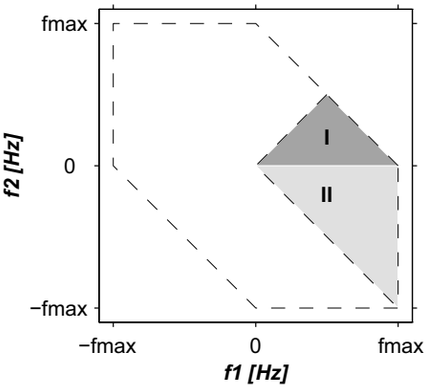


Fig. 4. Principal domain of the bicoherence. The Nyquist theorem restricts the display to the area enclosed by a dashed line. For the autobicoherence, the principal domain is I, for the cross-bicoherence it consists of I and II.

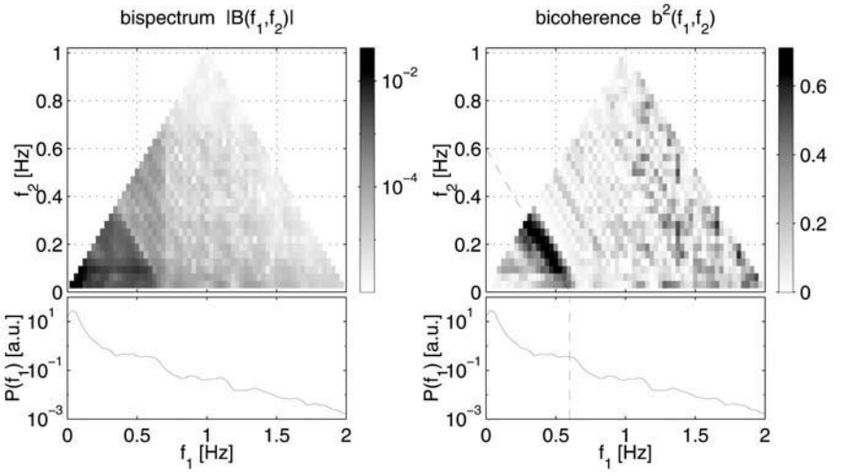


Fig. 5. Autobispectrum (left) and autobicoherence (right) of the AMPTE magnetic field data. The magnetic field component with the largest variance is analysed and the display is restricted to the principal domain. For convenience the power spectral density is shown below each plot. The number of samples is 12000 and the display has deliberately been cut off at 2 Hz (the Nyquist frequency is 8 Hz).

Figure 5 shows the bispectrum and the bicoherence. The former reveals a featureless but significant level of wave coupling, with most of the energy concentrated in low frequency modes. The coupling strength, however, cannot be assessed by lack of normalization. The bicoherence in contrast shows a significant phase coupling between Fourier modes that satisfy the resonance condition $f_1 + f_2 = 0.6\,\mathrm{Hz}$, with $f_1 \leq f_2 \leq 0.6\,\mathrm{Hz}$. Such a phase coupling

suggests that the large-scale magnetic structures are somehow coupled to the whistler waves, whose frequency peaks around 0.6 Hz. The coupling is significant ($b^2 \approx 0.7$) but it does not reach higher levels because the nonlinear structures are embedded in a randomly fluctuating wavefield.

Example 2: A different aspect is revealed by water level fluctuations as recorded in a laboratory experiment. The swell results in regular and fairly monochromatic gravity waves on top of which small amplitude capillary waves are produced by wind. The bicoherence analysis of these water waves is shown in Fig. 6.

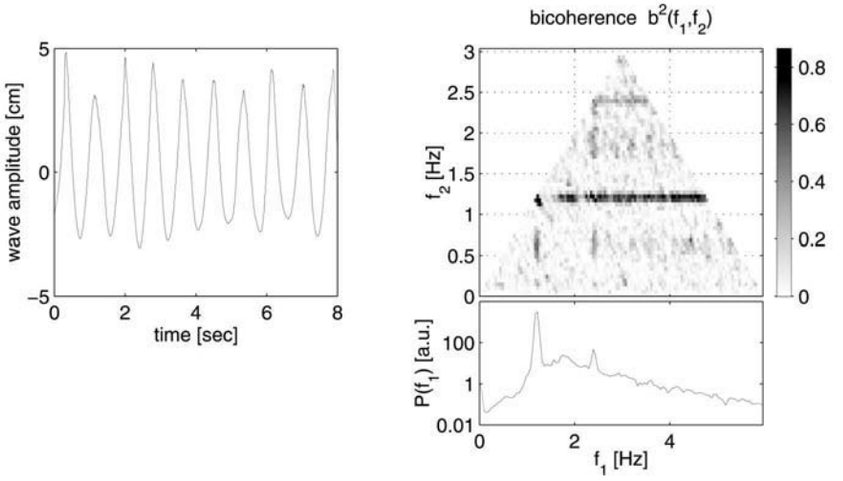


Fig. 6. Excerpt (left) and autobicoherence (right) of the water level fluctuations. For convenience the power spectral density is shown below the bicoherence. The number of samples is 65536 and the display has been cut off at 6 Hz.

The salient features of the bicoherence are a marked peak at $1.2 + 1.2 = 2.4$ Hz and a ridge $1.2 + f = 1.2 + f$ with $f > 1.2$ Hz. The former is the classical signature of harmonic generation (the fundamental is coupled to the first harmonic). One can also distinguish higher harmonics. The ridge means that the fundamental mode is phase coupled to all high-frequency modes, and not just to its harmonics. Indeed, the high frequency capillary waves are always located on top of the swell (i.e. they are phase coupled to the fundamental) because this is where they are most efficiently generated by the wind.

These two examples reveal the existence of significant phase couplings between specific Fourier modes. We must stress, however, that a phase coupling does not necessarily imply the existence of nonlinear wave interactions per se. In the first example, the ridge could be interpreted both as a decay ($0.6 \rightarrow f_1 + f_2$) or as an inverse decay ($f_1 + f_2 \rightarrow 0.6$) process. At this stage we cannot tell whether the observed phase coupling is accompanied by an energy transfer be-

tween Fourier modes (i.e. the wavefield is dynamically evolving) or whether it is just the remnant of some nonlinear effect that took place in the past or maybe even some nonlinear instrumental effect. This caveat has been highlighted by Pécseli and Trulsen [57]. To unambiguously assess nonlinear wave interactions, multipoint measurements are needed, see Sect. 2.6.

2.4 Estimating Higher Order Spectra

Higher order spectra can be estimated either by direct computation of the higher order moments from Fourier transforms, or by fitting the data with a parametric model.

The **Fourier approach** is computationally straightforward: the time series is divided into M sequences, for each of which the Fourier transform is computed. An unbiased estimate of the bispectrum is then

$$\hat{B}(\omega_k, \omega_l) = \frac{1}{M} \sum_{i=1}^{M} u_k^{(i)} u_l^{(i)} u_{k+l}^{*\,(i)} \ , \tag{13}$$

The empirical estimate of the bicoherence becomes

$$\hat{b}^2(\omega_k, \omega_l) = \frac{\left| \hat{B}(\omega_k, \omega_l) \right|^2}{\frac{1}{M} \sum_{i=1}^{M} \left| u_k^{(i)} u_l^{(i)} \right|^2 \frac{1}{M} \sum_{i=1}^{M} \left| u_{k+l}^{(i)} \right|^2} \ . \tag{14}$$

Careful validation of higher order quantities is essential as these quantities are prone to errors. Hinich and Clay [31] have shown that the variance of the bicoherence is approximately

$$\mathsf{Var}[\hat{b}^2] \approx \frac{4\hat{b}^2}{M} \left(1 - \hat{b}^2 \right) \ , \tag{15}$$

and that this quantity has a bias

$$\mathsf{bias}[\hat{b}^2] \approx \frac{4\sqrt{3}}{M} \ . \tag{16}$$

It is therefore essential to work with long and stationary time series (i.e. $M \gg 1$) in order to properly assess low bicoherence levels. This constraint becomes even more stringent for tricoherence estimates.

The need for long time series can be partly alleviated by using alternative spectral representations. Morlet wavelets, because of their better time-frequency resolution, can improve the estimates [74,20].

A different, so-called **parametric approach** consists in fitting the time series with a parametric model, such as an autoregressive (AR) model [43]. If this model correctly fits the data, the one can retrieve from it not only the spectrum but also higher order spectra. For a detailed description, see [54] and [48].

Parametric methods are appropriate for short time series, or for time series whose spectrum and higher order spectra are either featureless (i.e. they are smooth functions of the frequency), or contain a few narrow spectral lines [58]. Fourier methods are easier to apply and do not require any decision on the type of model that should be fitted.

2.5 Higher Order Spectra: More Properties

Higher order spectra share some other properties that can be of interest for the analysis of wave phenomena. The bicoherence, for example, can be used to detect asymmetries in a time series. Time-reversal asymmetries ($u(t) \leftrightarrow u(-t)$) give rise to purely imaginary bispectra, so to detect them one should use a variant of the bicoherence

$$b_i^2(\omega_k, \omega_l) = \frac{|\Im B(\omega_k, \omega_l)|^2}{\langle |u_k u_l|^2 \rangle \langle |u_{k+l}|^2 \rangle} \ . \tag{17}$$

Up-down asymmetries ($u(t) \leftrightarrow -u(t)$) show up in the real bispectrum only, so the quantity to be used is

$$b_r^2(\omega_k, \omega_l) = \frac{|\Re B(\omega_k, \omega_l)|^2}{\langle |u_k u_l|^2 \rangle \langle |u_{k+l}|^2 \rangle} \ . \tag{18}$$

The distinction between symmetries is useful for separating concurrent nonlinear processes: time-reversal asymmetries arise during nonlinear wave steepening, whereas up-down asymmetries may occur in wavefields with cavitons. These properties are illustrated in Fig. 7, which shows the "real" and "imaginary" bicoherences for the gravity wave example. Water waves are known to have strong up-down asymmetries, and this indeed shows up in large values of the "real" bicoherence b_r^2. The small value of b_i^2 suggests that the waves are still far from overturning, in spite of the wind strength.

Higher order spectra can easily be generalized from temporal to spatio-temporal couplings. Space plasmas are truly spatio-temporal systems, so a nonlinear wave coupling should not only involve resonant frequencies, but also resonant wavenumbers. Three-wave interactions, for example, can only occur between waves that satisfy the two resonance conditions

$$\begin{cases} \omega_k + \omega_l = \omega_p \\ \boldsymbol{k}_k + \boldsymbol{k}_l = \boldsymbol{k}_p \end{cases} \ . \tag{19}$$

The first condition can be interpreted as a conservation of energy while the second is a conservation in momentum. This resonance is illustrated in Fig. 8 for one-dimensional case.

In practice, analyses in both ω and $\boldsymbol{k}$ space are hard to achieve unless the data offer sufficient spatial and temporal coverage. The user thus often has to make hypotheses on one of the two quantities in order to compensate for the lack of information on the other. In the case of strong convection, or if the Taylor hypothesis (i.e. if the turbulence is frozen in the wavefield, see [26]) holds, then $\omega \propto |\boldsymbol{k}|$ and so no spatial resolution is actually needed.

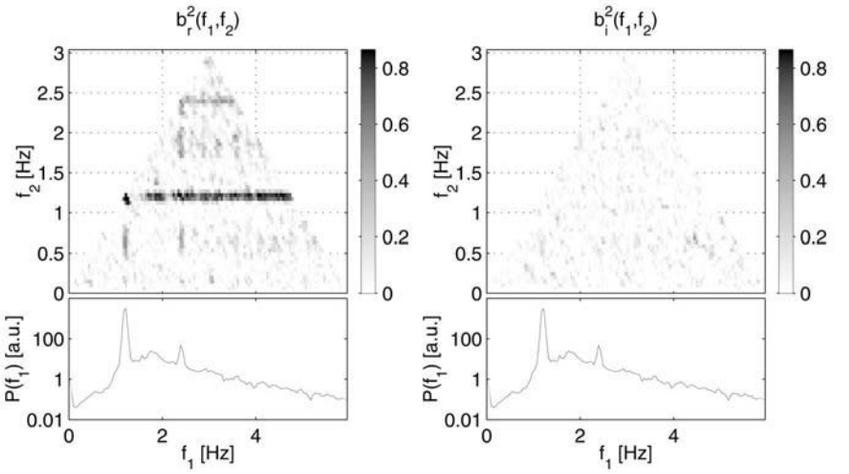


Fig. 7. Autobicoherence of the water level fluctuations (as in Fig. 6), using the real bispectrum (left) and the imaginary one (right).

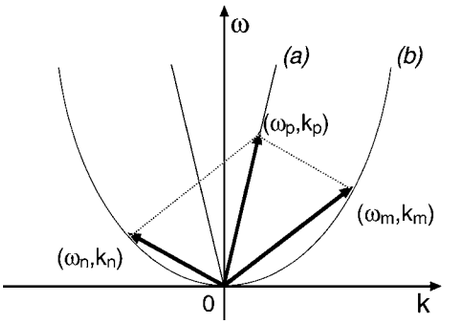


Fig. 8. Example of a plasma with two dispersion relations (a) and (b), and a possible three-wave interaction, described as vectors in (k, ω) space. Only one dimension in space is considered here.

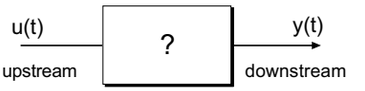


Fig. 9. The plasma behaves like a black box, which reacts to the input $u(t)$ by giving a response $y(t)$. The nonlinear properties of the plasma are recovered by modelling this response.

2.6 Spectral Energy Transfers

One of the main shortcomings of higher order spectra is their inability to tell the origin of the phase coupling. Is it the consequence of an instrumental nonlinearity, is it the remnant of a nonlinearity that took place some time during the wavefield

evolution, or does it result from an ongoing dynamical process? To answer this question, we must go back to the Volterra kernels (eq. 3) and define a new quantity, called *spectral energy transfer* [69,67].

For a large ensemble of waves with different frequencies, the random phase approximation implies that

$$\langle u_p^* u_q \rangle = P_p \, \delta_{pq} \ , \tag{20}$$

where $P_p = P(\omega_p)$ is the power spectrum. Combining this with eq. 4 gives

$$\frac{\partial P_p}{\partial x} = \gamma_p P_p + \sum_{m+n=p} T_{m,n} + \sum_{m+n+k=p} T_{m,n,k} + \cdots \tag{21}$$

This equation expresses the spatial variation in the spectral energy density at a given frequency (or wavenumber) as a sum of linear and nonlinear terms. The first term ($\gamma_p P_p$) describes linear processes, the second ($T_{m,n}$) quadratic processes, etc. The latter tells us how much energy is being transferred to ($T > 0$) or away ($T < 0$) from the p'th Fourier mode through nonlinear interactions with other modes [35].

Spectral energy transfers are among the most relevant quantities for describing nonlinear wave interactions. Unfortunately, they are rarely estimated because they require high spatio-temporal coverage and involve ill-conditioned systems. Significant results have nevertheless been reported with quadratically nonlinear models in laboratory plasma turbulence [69], for magnetospheric turbulence [23], and for Langmuir wave turbulence [72].

The numerical problems encountered in estimating spectral energy transfers may be alleviated by estimating the Volterra kernels in the time domain, using parametric models, and subsequently computing the energy transfers in Fourier space. Indeed, time-domain models are often quite efficient in capturing nonlinear features in a small number of terms. Polynomial models offer flexibility at a reasonable computational expense. The most widely used class of models is called Nonlinear AutoRegressive Moving Average with eXogeneous input (NARMAX) [56].

Consider the simplest case in which two measurements $y(t)$ and $u(t)$ of the same quantity are made simultaneously at closely spaced locations. Let the time evolution of one be a direct consequence of the other. This typically occurs when making two-point measurements in a convected wavefield. The nonlinear transfer function which relates the "output" $y(t)$ to the "input" $u(t)$ then embodies the nonlinear dynamics of the plasma, which would be called a black box. By modelling this transfer function, one can recover the salient nonlinear properties of the plasma, as shown by Ritz and Powers [67]. This concept not only applies to turbulence, but also to solar-terrestrial interactions, where it can be used to make predictions [76].

In practice, both $u(t)$ and $y(t)$ are continuously sampled, so a discrete transfer function model is needed. NARMAX models express the output $y_i = y(t_i)$ at a given time as a polynomial involving all possible combinations (up to a given

order) of past outputs, past and present inputs, and past residual errors

$$y_i = \mathcal{P}[y_{i-1}, \ldots, y_{i-n}, u_i, u_{i-1}, \ldots, u_{i-n}, \varepsilon_{i-1}, \ldots, \varepsilon_{i-n}] + \varepsilon_i \ , \tag{22}$$

where the residual error ε_i is the difference between the true output and the model response.

An application of NARMAX modelling to magnetospheric turbulence (based on the AMPTE dataset of this book) was reported in [17]. The capacity of such models to fit nonlinear systems with a few polynomial terms is clear asset. Unfortunately some of the mathematical tools associated with these models are not in the public domain.

3 Higher Order Statistics

Higher order statistics and higher order spectra are intimately related, but they are generally encountered in different contexts. When a system is driven away from the linear-deterministic corner in Fig. 1, then time-domain representations are often more appropriate than spectral representations.

Most of the literature on probability theory and statistics deals with random variables whose probability density[5] function (pdf) is a Gaussian. Gaussian distributions, however, are not all there is to statistics. Very often it is precisely the deviation from Gaussianity that contains pertinent information about the underlying physics. The tails of the distribution, which represent large but rare events, are of particular interest and have received much attention. Many statistical tools have been developed for this purpose. Most of them have emerged in the context of turbulence [26], but new paradigms such as complexity [5], self-organization [34] and anomalous transport [10] today stimulate the development of novel tools. Meanwhile, it should be stressed that the proper estimation of a pdf itself can already be a tricky task; see [70] for a thorough study.

3.1 Scale Invariance

Symmetry is one of the key concepts behind higher order statistical analysis. Many physical systems exhibit symmetries. In turbulent wavefields, for example, quantities like the velocity field often remain unchanged under the following transformation

$$\mathbf{v}(\mathbf{x}, t) \longrightarrow \lambda \mathbf{v}(\lambda^a \mathbf{x}, \lambda^b t) \tag{23}$$

Because of these symmetries, the system is said to be *scale invariant*. Roughly speaking, scale invariance means that within a wide range of (spatial or temporal) scales, it is not possible to identify a predominant scale. The property of interest is the interplay between scales, rather than the role played by characteristic scales. This property is typical of thermodynamical systems that are at a critical point and can be quantified in many different ways.

[5] From a mathematical point of view, the terms *probability density* and *probability distribution* bear different meanings. Here, as in the physics literature, they are used interchangeably.

Spectral Signatures. Scale invariance can easily be expressed in the Fourier domain. Equation 23 implies that the Fourier spectra in frequency (and in wavenumber) of a scale invariant quantity behave like power laws with no characteristic cutoff frequency (respectively cutoff wavenumber)

$$P(\omega) \propto \omega^{-\beta} \ . \tag{24}$$

It should be stressed that this property is a necessary but not a sufficient condition. The same power law scaling should hold for all higher order spectra. Structure function analysis (see below) will provide a means for getting more information.

Exact power laws never occur in real life. Mesoscales always provide some low-frequency cutoff (otherwise the variance of the process would diverge) whereas damping or viscosity cause a high frequency cutoff. In between, the spectrum is often corrupted by spectral lines associated with plasma waves. In spite of this, power laws extending over several decades have been observed in the solar wind [29] and in simulations. Inside the magnetosphere, this range is generally shorter. For a power law to be meaningful, it should cover at least one decade.

Equation 24 suggests that the scaling exponent or spectral index β can be directly estimated from the Fourier power spectrum. It has been shown since [3] that wavelets are better for that purpose because they are inherently self-similar. The distinction between the two approaches becomes particularly important when one has to deal with short and noise-corrupted data. Figure 10 shows an example based on a synthetic data set. A time series with $N = 512$ samples was generated with a spectral index $\beta = 2$. The figure compares the spectrum obtained by standard FFT and by using a discrete wavelet transform with Daubechies wavelets. Clearly, the latter succeeds much better in capturing the power law scaling.

Scale-Invariance and Structure Functions. Since the spectrum is not sufficient for assessing scale invariance, it is necessary to look at higher order moments. In addition to this, one should also consider spatial gradients and higher order derivatives in order to distinguish temporal variations from spatial structures. For a stationary Gaussian process, these derivatives all must have a Gaussian pdf.

In practice, it is quite difficult to properly separate spatial structures from temporal variations, even with multipoint measurements. Therefore, instead of computing gradients, it is customary to compute spatial increments

$$\delta\mathbf{y} = \mathbf{y}(\mathbf{x} + \mathbf{l}, t) - \mathbf{y}(\mathbf{x}, t) \ . \tag{25}$$

Furthermore, assumptions are often made to convert spatial structures into temporal dynamics. In the solar wind, for example, the turbulence is nearly frozen in the wavefield (this is better known as the Taylor hypothesis, see [26]) so that one can reasonably set $\tau = l/v$ and hence for a scalar quantity

$$\delta y = y(t + \tau) - y(t) \ . \tag{26}$$

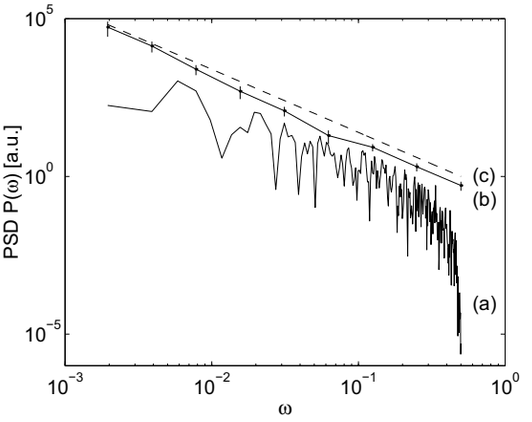


Fig. 10. Power spectral density of a synthetic time series with $N = 512$ samples and a spectral index $\beta = 2$: a) stands for standard periodogram estimate with a single Welch window, b) was obtained with 3rd order Daubechies wavelets (the scales a are converted into frequencies by $\omega = 1/a$), c) gives the exact result. The power spectra have been shifted vertically for easier comparison.

Figure 11 shows an excerpt of magnetic field data gathered by Ulysses in 1994 within the solar wind from the south solar pole. The wavefield mainly consists of Alfvén waves that rotate on a sphere of constant radius, hence the peculiar pdf, which is much more akin to a uniform than to a Gaussian distribution. The pdfs of the differenced data on the contrary exhibit strongly enhanced tails. For small τ, the time series shows large bursts and the pdf is highly non-Gaussian. The larger τ is, the closer the pdfs are to the original one, because distant fluctuations are essentially independent. The study of this departure from Gaussianity is a central research issue in many fields because it is indicative of the microscopic turbulent processes.

The simplest way of quantifying the various shapes of the pdfs consists is to compute their higher order moments. This is called the *structure function* approach

$$S_p(\tau) = \langle |y(t+\tau) - y(t)|^p \rangle \ . \tag{27}$$

A large amount of literature has been devoted to this technique. For general references, see [55,26,9]; introductions to plasmas can be found in [8,47,64]. Structure functions can in principle be computed from any physical quantity, but a meaningful comparison against theory requires the use of natural variables. In space plasmas, the natural variables for Alfvén waves are the Elsässer variables $\mathbf{Z}^{\pm} = \mathbf{V} \pm \mathbf{V}_A$, where $\mathbf{V}_A$ is the Alfvén velocity [47].

An interesting result is the existence of a universal scaling law when the wavefield is scale-invariant

$$S_p(\tau) \propto \tau^{\alpha(p)}. \tag{28}$$

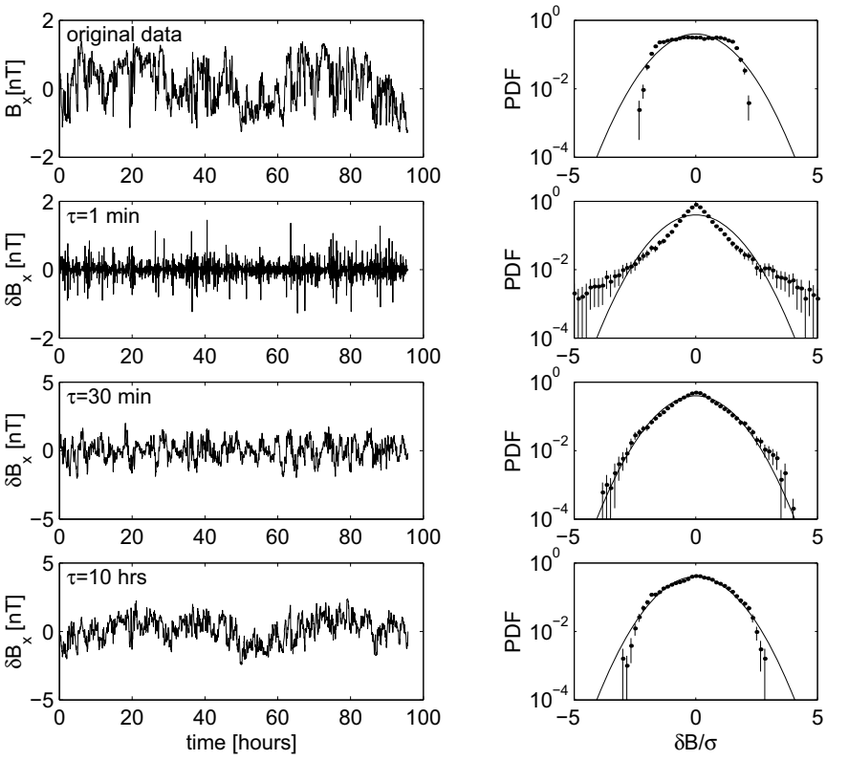


Fig. 11. Excerpt of Ulysses magnetic field data (left column) and the associated pdfs (right column). Each row corresponds to a different value of τ, the first row showing the original data. The number of samples is 20000 and the B_x component was used. All the pdfs have been rescaled to have unit standard deviation and zero mean; a Gaussian distribution with zero mean and unit standard deviation is shown for comparison. The sampling period is 1 min. and error bars correspond to ± 1 standard deviation.

The larger order p is, the more emphasis is put on the tails of the distribution and the more difficult it becomes to measure the scaling exponent $\alpha(p)$. The range in which this scaling holds is called inertial range. Any deviation of $\alpha(p)$ from a linear dependence is an indication for irregular redistribution of the energy in the turbulent cascade. In the classical K41 model by Kolmogorov [26], the turbulent eddies are supposed to be in a state of local equilibrium: each eddy decays into smaller ones, which again give rise to smaller eddies, and so on. For this model, one obtains $\alpha = p/3$. Many theoretical models for turbulent cascades have since been developed to match the observations [55,9]. The model developed by Castaing and coworkers is today widely used both in neutral fluids and in plasmas [64].

As an example, consider a turbulent cascade in which the eddies are not space filling but occupy a domain whose physical dimension is $D \leq 3$. One

should then have $\alpha(p) = 3 - D + p(D-2)/3$. This property is illustrated in Fig. 12 for three cases: a) the K41 model in which each eddy is space-filling (D=3), b) the *monofractal* case in which each eddy occupies a fixed fraction of space (D is fixed), and c) the *multifractal* case in which the occupancy of space varies locally (D varies). A sparse filling of space typically results from a stretching of eddies into filamentary vortices. Numerical simulations indeed support such a picture [8].

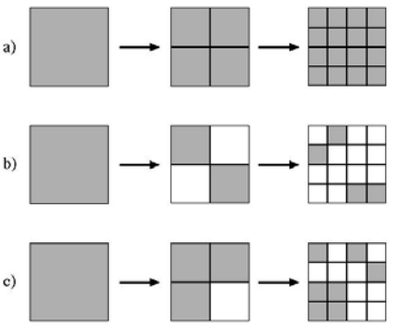


Fig. 12. Illustration of the turbulent cascade in 2D, in which each turbulent eddy (gray area) decays into four smaller eddies. Cases a) to c) respectively correspond to the the Kolmogorov K41 model with space-filling eddies, the monofractal cascade model, and the multifractal cascade model.

Burlaga [12] first pointed out the striking similarity between solar wind turbulence and hydrodynamic turbulence, a result confirmed since by many [46,13,32]. Structure function analysis has also been applied to other quantities, such as geomagnetic indices [73]. All these studies suggest that space plasmas are scale-invariant but that this scaling varies locally, hence the name multifractal. The macroscopic consequence of this local property is irregular behaviour with sudden bursts of activity called *intermittency*.

The structure function associated with the Ulysses data is shown in Figure 13, for orders between $p = 1$ and $p = 6$. It can be argued that an inertial range is apparent for $\tau = 3 - 30$ min. In this range, the scaling exponent is a convex function of p, which supports the multifractal character of the turbulent wavefield. The figure also shows another quantity

$$A_p(\tau) = \frac{S_p(\tau)}{S_2(\tau)^{p/2}} \ . \tag{29}$$

For a Gaussian distribution, $A_p(\tau)$ should not depend on the scale τ. This quantity confirms the existence of wide distributions for small τ and on the contrary rather narrow distributions for large τ.

Structure Functions: Pitfalls. Structure functions suffer from a number of problems. First, theoretical models are not as easy to develop for magnetofluids

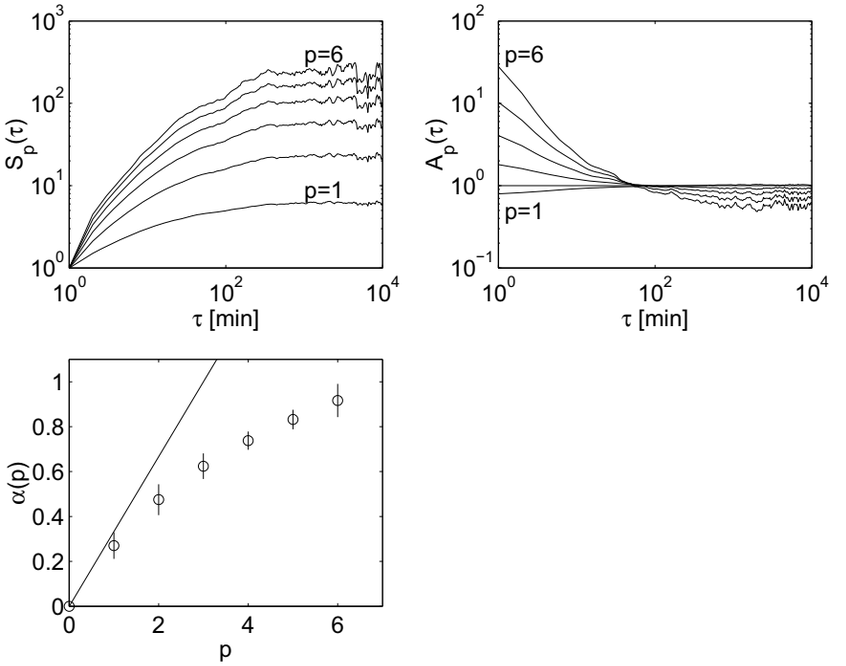


Fig. 13. Structure function analysis of Ulysses magnetic field data: the structure function $S_p(\tau)$ (top left), the normalized structure function $A_p(\tau)$ (top right), and the scaling exponent $\alpha(p)$ estimated between $\tau = 3$ and $\tau = 30$ (bottom left). The data set is the same as in Fig. 11.

as for neutral fluids. Some obstacles to a straightforward interpretation are symmetry breaking due to the magnetic field, the need to have Elsässer variables, and the questionable validity of the Taylor hypothesis in space plasmas.

Furthermore, structure functions, like all higher order quantities, become very vulnerable to outliers and lack of statistics as the order p increases. This problem has often been overlooked in the literature. It turns out that moments beyond $p = 4$ or $p = 5$ often cannot be meaningfully assessed, even with large data sets [21,32]. Therefore, great care should be taken in interpreting results.

Scale Invariance: Beyond Structure Functions. Structure functions have enjoyed great popularity so far, but there are many alternatives to the characterization of scale invariance. It has been shown [51], for example, that wavelet-based structure function estimates perform better. The reason for this is the self-similar nature of wavelets.

Scale invariance can be probed by many other quantities, which are conceptually related the topological invariants, information measures, etc. Since the domain of application largely exceeds plasma physics and nonlinear time series analysis, any classification tends to be arbitrary. We just mention here

multifractal analysis, whose purpose is to measure the *singularity spectrum* (or multifractal spectrum). The latter is indicative of the dimension of the region over which a spectral index β of a given value is observed. Singularity spectra are in turn connected to generalized dimensions [5]. Figure 14 shows an example of singularity spectrum that was computed by Consolini [18] from the auroral electrojet (AE) geomagnetic index. For a monofractal process, only one value of β should occur, whereas we observe a broad distribution of values that indicative of a multifractal process.

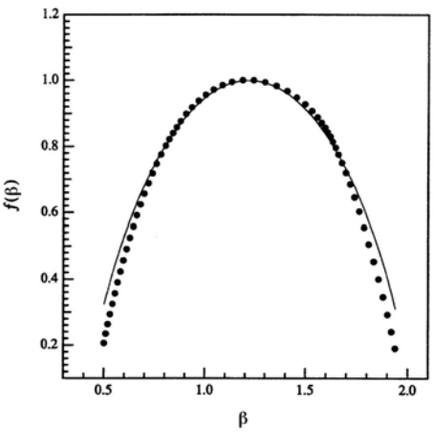


Fig. 14. Singularity spectrum from auroral electrojet data. Dots are experimental data points and the line comes from a model that is based on a two-scale Cantor set. Figure adapted from [18].

The robust estimation of singularity spectra is a delicate task and validation is not straightforward. The estimation of singularity spectra can be done in various ways, e.g. [16]. These techniques are outperformed today by a wavelet method called *wave transform modulus maxima* [4]. Clearly, the prime result of interest is not necessarily the singularity spectrum itself, but its variation under different conditions. There remains an important issue to determine under what conditions it remains invariant.

Finally, let us mention the theory of large deviations [71] as an alternative approach to the modelling of pdfs with long tails. Its relevance for many geophysical phenomena (such as earthquakes) may well extend to space plasmas.

3.2 Long-Range Dependence and Self-organized Criticality

Long-range dependence is just another manifestation of scale invariance. But, since it is often studied in connection with (Forced) Self-Organized Criticality (FSOC or SOC), it deserves a section on its own.

The SOC paradigm has recently enjoyed great popularity as a possible explanation for a wide range of phenomena that are observed in complex systems. SOC occurs in spatially-extended metastable systems, in which small perturbations can trigger fluctuations or avalanches of any size. The main signatures of SOC are scale invariance and long-range correlations, and a fractal topology (i.e. a self-similar spatial structure). The existence of long-range correlations, however, does not imply the existence of SOC per se. For a general reference on SOC, see [34] and for a discussion on the quantification of long-range dependence see [7]. Whether SOC is really applicable to space plasma phenomena such as turbulence is still a matter of controversy (e.g. [14]); meanwhile, some fingerprints of SOC have been identified in magnetospheric dynamics [77,19].

Spectral Signatures. Processes with long-range dependence are sometimes termed $1/f$ processes because they often exhibit power law spectra $P(\omega) \propto \omega^{-\beta}$ with a spectral index β that is close to 1. The limiting case $\beta = 1$ corresponds to a singularity with infinite range correlations. Therefore, a necessary (but not a sufficient) condition for having long-range dependence is to observe spectra that follow a power law down to the smallest frequencies (or wavenumbers). A cutoff always eventually occurs due to the finite size of the system.

As argued before in Sec. 3.1, the estimation of the spectral index β should be done with great care. In particular, estimators based on wavelets [3] should systematically be preferred.

Figure 15 shows an application to the AE geomagnetic index, a quantity that is often used because of its close connection with the dynamics of substorms. One year of one-minute resolution data were analyzed. The Fourier spectrum reveals a broken power law, with approximately a ω^{-2} scaling at high frequencies and a ω^{-1} scaling below. The wavelet-based spectrum gives much better evidence for this broken power law, and in addition reveals the cutoff frequencies more evidently. The spectrum eventually saturates around a period of a few tens of days, which is likely to be associated with the solar rotation period. From the wavelet spectrum, we estimate the following spectral indices: $\beta = 0.96 \pm 0.18$ for periods from about 4 hours up to 4 days, and $\beta = 1.87 \pm 0.03$ for periods between 1 minute and approximately 4 hours. An accurate assessment of these spectral indices is important for determining the underlying physical models.

Autocorrelation. Most simple physical models (i.e. Markov models, autoregressive models) give rise to exponentially decaying autocorrelation functions. Such a functional dependence means that the process has a characteristic time scale (or spatial scale). For scale invariant processes, the autocorrelation function should instead decay algebraically

$$C(\tau) \propto \tau^{-\gamma} \ , \tag{30}$$

where the exponent is connected to the spectral index by $\gamma = 1 - \beta$. Strictly speaking, a process is called long-range correlated when $0 < \gamma < 1$.

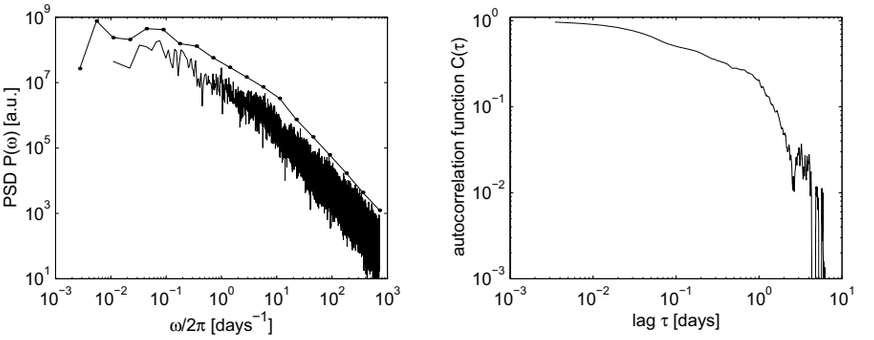


Fig. 15. Power spectrum (left) and autocorrelation function (right) of one year of AE data from 1986. The sampling period is 1 minute. The power spectrum was estimated using a periodogram with 5 windows, and the wavelet spectrum (dots) with 4th order Daubechies wavelets. The wavelet scales are connected to frequencies by $\omega = 1/a$. The two spectra have been shifted vertically to ease comparison.

Unfortunately, autocorrelation functions provide a poor estimate of the indices β and γ [7]. This is illustrated in Fig. 15: note how difficult it is to recognize a power law scaling in the autocorrelation function; an exponential gives almost as good a fit.

Waiting Time Statistics. Another interesting statistical measure is the waiting-time or inter-burst time distributions. The idea is to consider the interval between bursts rather than the burst size. Many simple models yield distributions of intervals that follow a Poisson law with exponential tails, whereas scale invariant systems should give a power law. See [25] for an application to the AE index. These methods are receiving today much attention in the context of solar physics [2].

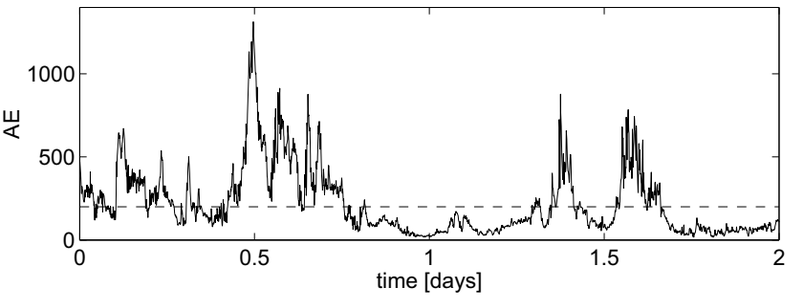


Fig. 16. Excerpt of the AE index data. Waiting times are based on the intervals between each successive crossing of the $AE = 200$ line.

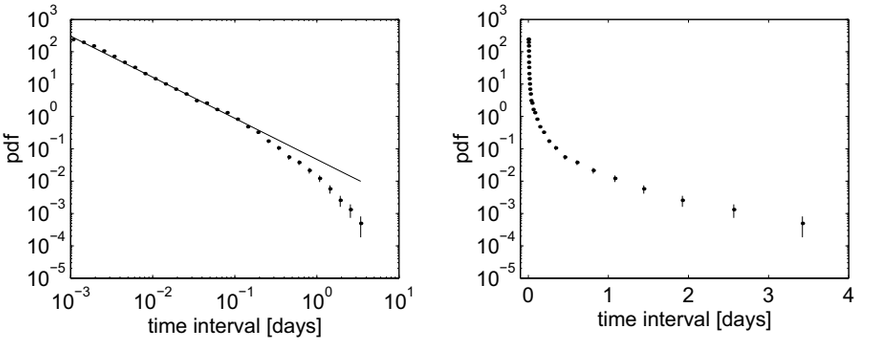


Fig. 17. Waiting time statistics for the AE index in 1986. The pdfs of the time intervals between successive crossings of the $AE = 200$ amplitude are shown. Left plot is with log-log axes and plot with lin-log axes.

An example is shown in Fig. 17, again based on one year of AE index data. An excerpt of the time series is shown in Fig. 16. We computed the duration of the intervals between successive values where the index crosses the $AE = 200$ amplitude line. Changing this level does not significantly affect the results as long as it stays in the bulk of the distribution. The pdf of the time intervals clearly shows a power law scaling that extends over almost two decades (left plot), and which can definitely not be fitted with an exponential (right plot). For small intervals, the results are indicative of the magnetospheric dynamics, whereas long intervals may be affected by external driving terms such as the solar wind, hence the deviation from the power law scaling.

Other Measures. One can think of many other possible measures for long-range dependence; the main problem is to keep physical insight. The *Hurst exponent* has been popular lately in fusion plasmas, even though its interpretation can be ambiguous [7]. The method itself is called *Hurst rescaled range analysis*; it was originally proposed by Hurst, who used it to detect persisting trends in time series (see for example [45]). A recent application of the Hurst exponent to tokamak plasma turbulence was used to claim the existence of long-range dependence and SOC [15]. Similar conclusions were drawn from the analysis of AE data [60].

3.3 Lévy Walks and Anomalous Transport

Recently, a great deal of attention has been paid to anomalous (i.e. non-diffusive) transport in fluids. For low frequency magnetic turbulence and a strong background magnetic field, the motion of charged particles is approximately along the magnetic field lines. When the fluctuation level increases and/or the anisotropy of the magnetic field changes, the braided topology of the magnetic field may generate new types of transport regimes that are not necessarily diffusive anymore. This again results in scale-invariance and non-Gaussian properties.

Anomalous transport can be appropriately described in the framework of Lévy distributions [10,62]. Consider the probability distribution of the distance and the duration of a particle trajectory between two successive bounces. Anomalous transport is intimately associated with distributions that asymptotically decay as power laws. This means that the particle has a significant probability to move long distances. Furthermore, some of the moments of the distribution (such as the mean distance, the variance, etc.) may actually diverge. Studies of Lagrangian transport of tracer particles in fluid experiments [63] and in plasma turbulence simulations [59] have revealed the mechanisms by which Lévy walks occur. The resulting anomalous transport can have an noticeable impact on macroscopic plasma properties, and may explain for example percolation from the solar wind into the magnetosheath.

Fig. 18. Trajectory in 2D space of a particle with Brownian motion (left) and Lévy random walk (right).

Figure 18 compares the trajectories in 2D space of a particle with standard random walk motion (i.e. Brownian motion) and a particle that makes Lévy random walks. The first has a diffusive motion whereas the second will give super-diffusion. The characterization of such transport regimes is appropriately done by using a Lagrangian approach with test particles. Let $\langle \Delta x^2 \rangle_\tau$ be the mean squared distance traveled by a particle during a time interval τ. Then

$$\langle \Delta x^2 \rangle_\tau = 2D\tau^\mu \ ,$$

where D is the diffusion coefficient and μ the diffusion exponent. Brownian motion with standard diffusion corresponds to the case $\mu = 1$, $\mu < 1$ is called subdiffusion, $\mu > 1$ superdiffusion, and $\mu = 2$ corresponds to ballistic motion[6].

As an example we consider test particle simulations based on a 2-D model [39]. The scaling of the mean squared displacement with the observation time is shown in Fig. 19. A power law is indeed apparent over several decades. Below the ion gyrofrequency, the scaling exponent is $\alpha = 2$, in agreement with the ballistic nature of the ion gyration. Above the gyrofrequency, however, the exponent α

[6] Note that for processes with Gaussian statistics, the value of μ is connected to the spectral index β.

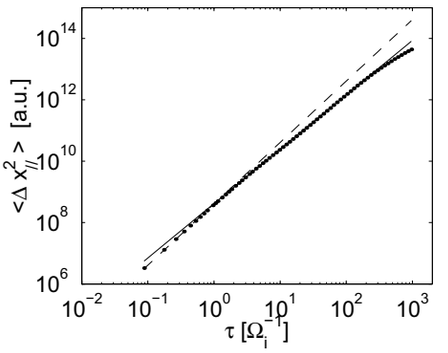


Fig. 19. Scaling of the mean square displacement along the magnetic field vs travel time τ. The slopes of the straight lines are $\mu = 1.8$ (full line) and $\mu = 2.0$ (dashed line).

stays close to 1.8, thereby suggesting that the motion is neither diffusive, nor ballistic. The deviation at largest time intervals is a finite sample size effect.

4 Outlook

The development of techniques for characterizing nonlinear processes is a challenging and rapidly expanding area. However, as the techniques tend to become more and more sophisticated, there appears the risk of losing physical insight. It is indeed tempting to gather many techniques in a toolbox, and then start doing data mining.

In the next decades, significant outbreaks are expected (or at least hoped for) in the following areas:

- The analysis of spatio-temporal processes, which include data from multipoint measurements, 1D, 2D and 3D models. Our capacity of analyzing data dramatically drops as soon as a spatial dimension comes in, and most of the results obtained so far are still based on time-domain techniques.
- As we increasingly deal with large and multivariate data sets that contain a lot of redundant information, there is also a growing need for doing preprocessing. This involves "reducing" the number of significant variables to make them more tractable. It also means locating interesting features in the data. Many techniques have been developed for that purpose, using either classical methods (e.g. principal component analysis) or more novel concepts such as artificial intelligence.
- Models have a significant advantage over experiments: they can produce long records of data with a very low noise level. Many nonlinear models in addition show signatures of low-dimensional determinism, and therefore lend themselves for a characterization of chaotic behaviour. Tools that have been developed for that purpose (e.g. [61]) can nowadays provide deep insight into the analytical properties of such models.

References

1. H. D. I. Abarbanel, R. Brown, and J. J. Sidorovich: *The analysis of observed chaotic data in physical systems*, Rev. Mod. Phys. 65, 1331-1392 (1993).
2. M. J. Aschwanden, B. R. Dennis, and A. Benz: *Logistic Avalanche Processes, Elementary Time Structures, and Frequency Distributions in Solar Flares*, Astrophysical Journal 497, 972-993 (1998).
3. P. Abry, P. Gonçalvès, and P. Flandrin: *Wavelets, spectrum estimation,* $1/f$ *processes*, in *Wavelets and Statistics*, edited by A. Antoniadis and G. Oppenheim, Lecture Notes in Statistics, Vol. 103 (Springer, Berlin, 1995), pp. 15-30.
4. E. Bacry, J.F. Muzy, and A. Arneodo: *Singularity spectrum of fractal signals from wavelet analysis: exact results*, J. Stat. Phys. 70, 635-674 (1993).
5. R. Badii and A. Politi: *Complexity* (Cambridge Univ. Press, Cambridge, 1997).
6. S. D. Bale, D. Burgess, P. J. Kellogg, K. Goetz, R. L. Howard, and S. J. Monson:*Evidence of three-wave interactions in the upstream solar wind*, Geoph. Res. Lett., 23, 109-112 (1996).
7. J. Beran: *Statistics for long-memory processes*, (Chapman & Hall, London, 1994).
8. D. Biskamp: *Nonlinear magnetohydrodynamics*, (Cambridge Univ. Press, Cambridge, 1993).
9. T. Bohr, M. H. Jensen, G. Paladin, and A. Vulpiani: *Dynamical systems approach to turbulence*, (Cambridge Univ. Press, Cambridge, 1998).
10. J.-P. Bouchaud and A. Georges: *Anomalous diffusion in disordered media: statistical mechanisms, models, and physical applications*, Phys. Reports 195, 127-293 (1990).
11. D. R. Brillinger: *Time series*, (Holt, New York, 1975).
12. L. F. Burlaga: *Multifractal structure of the interplanetary magnetic field*, Geoph. Res. Lett. 18, 69-72 (1991).
13. V. Carbone: *Scaling exponents of the velocity structure functions in the interplanetary medium*, Ann. Geoph, 12, 585-590 (1994).
14. V. Carbone, et al.: *To what extent can dynamical models describe statistical features of turbulent flows?*, Europhys. Lett, 58(3), 349-355 (2002).
15. B. A. Carreras, B. P. van Milligen, M. A. Pedrosa, et al.: *Experimental evidence of long-range correlation and self-similarity in plasma fluctuations*, Phys. Plasmas 6, 1885-1892 (1999).
16. A. Chhabra and R. V. Jensen: *Direct determination of the* $f(\alpha)$ *singularity spectrum*, Phys. Rev. Lett. 62, 1327-1330 (1989).
17. D. Coca, M. A. Balikhin, S. A. Billings, H. St. C. K. Alleyne, M. W. Dunlop, and H. Lühr: *Time-domain identification of nonlinear processes in space-plasma turbulence using multi-point measurements*, in *Proc. Cluster II workshop on Multiscale/Multipoint Plasma Measurements,* document ESA SP-449, ESA, Paris (2000).
18. G. Consolini, M. F. Marcucci, and M. Candidi: *Multifractal structure of the auroral electrojet index data*, Phys. Rev. Lett., 76, 4082-4085 (1996).
19. G. Consolini and T. Chang: *Magnetic field topology and criticality in GEOTAIL dynamics: relevance to substorm phenomena*, Space Science Rev. 95, 309-321 (2001).
20. T. Dudok de Wit and V. V. Krasnosel'skikh: *Wavelet bicoherence analysis of strong plasma turbulence at the Earth's quasiparallel bow shock*, Phys. Plasmas 2, 4307-4311 (1995).
21. T. Dudok de Wit and V. V. Krasnosel'skikh: *Non-Gaussian statistics in space plasma turbulence: fractal properties and pitfalls*, Nonl. Proc. Geoph. 3, 262-273 (1996).

22. T. Dudok de Wit: *Data analysis techniques for resolving nonlinear processes in plasmas: a review*, in *The URSI Review on Radio Science 1993–1995*, edited by R. E. Stone (Oxford University Press, Oxford, 1996).
23. T. Dudok de Wit, V. Krasnosel'skikh, M. Dunlop, and H. Lühr: *Identifying nonlinear wave interactions using two-point measurements: a case study of SLAMS*, J. Geophys. Res. 104, 17079-17090 (1999).
24. P. Escoubet, A. Roux, F. Lefeuvre, and D. LeQuéau (eds.): *START: Spatio-temporal analysis for resolving plasma turbulence*, document WPP-047 (ESA, Paris, 1993).
25. M. P. Freeman, N. W. Watkins, and D. J. Riley: *Evidence for a solar wind origin of the power law burst lifetime distribution of the AE indices*, Geoph. Res. Lett. 27, 1087-1090 (2000).
26. U. Frisch: *Turbulence* (Cambrisge Univ. Press, Cambridge, 1995).
27. G. B. Giannakis and E. Serpedin: *A bibliography on nonlinear system identification*, Signal Processing 81, 533-580 (2001).
28. K.-H. Glassmeier, U. Motschmann, and R. Schmidt: *CLUSTER workshops on data analysis tools, physical measurements and mission-oriented theory*, document SP-371 (ESA, Paris, 1995).
29. M. L. Goldstein and D. A. Roberts: *Magnetoshydrodynamic turbulence in the solar wind*, Phys. Plasmas 6, 4154-4160 (1999).
30. P. Grassberger and I. Procaccia: *Characterization of strange attractors*, Phys. Rev. Lett. 50, 1265 (1983).
31. M. J. Hinich and C. S. Clay: *The application of the discrete Fourier transform in the estimation of power spectra, coherence and bispectra of geophysical data*, Rev. Geoph. 6, 347-363 (1968).
32. T. S. Horbury and A. Balogh: *Structure function measurements of the intermittent MHD turbulent cascade*, Nonl. Proc. Geoph. 4, 185-199 (1997).
33. W. Horton and A. Hasegawa: *Quasi-two-dimensional dynamics of plasmas and fluids*, Chaos 4, 227-251 (1994).
34. H. J. Jensen: *Self-organized criticality* (Cambridge University Press, Cambridge, 1998).
35. B. B. Kadomtsev: *Plasma turbulence* (Academic Press, New York, 1965).
36. H. Kantz and T. Schreiber: *Nonlinear time series analysis* (Cambridge University Press, Cambridge, 1997).
37. Y. C. Kim and E. J. Powers: *Digital bispectral analysis and its applications to nonlinear wave interactions*, IEEE Trans. Plasma Sci. PS-7, 120-131 (1979).
38. V. Kravtchenko-Berejnoi, F. Lefeuvre, V. V. Krasnosel'skikh, and D. Lagoutte: *On the use of tricoherent analysis to detect non-linear wave interactions*, Signal Processing 42, 291-309 (1995).
39. Y. Kuramitsu and T. Hada: *Acceleration of charged particles by large amplitude MHD waves: effect of wave spatial correlation*, Geophys. Res. Lett. 27, 629-632 (2000).
40. J.-L. Lacoume, P.-O. Amblard, and P. Comon: *Statistiques d'ordre supérieur pour le traitement du signal* (Masson, Paris, 1997).
41. D. Lagoutte, D., F. Lefeuvre, and J. Hanasz: *Application of bicoherence analysis in study of wave interactions in space plasmas*, J. Geoph. Res. 94, 435-442 (1989).
42. B. Lefèbvre, V. Krasnosel'skikh, and T. Dudok de Wit: *Stochastization of phases in four-wave interaction*, in P. G. L. Leach et al.: *Dynamical systems, plasmas and gravitation* (Springer Verlag, Berlin, 1999), pp. 117-129.
43. L. Ljung: *System identification: theory for the user* (Prentice-Hall, Englewood Cliffs, NJ, 1999).

44. J. L. Lumley: *Stochastic tools in turbulence* (Academic Press, 1970).
45. B. Mandelbrot: *Fractals and scaling in finance* (Springer Verlag, Berlin, 1997).
46. E. Marsch and C.-Y. Tu: *Non-Gaussian probability distributions of solar wind fluctuations*, Ann. Geoph. 12, 1127-1138 (1994).
47. E. Marsch: *Analysis of MHD turbulence: spectra of ideal invariants, structure functions and intermittency scalings*, in [28] pp. 107-118.
48. J. M. Mendel: *Tutorial on Higher Order Statistics (Spectra) in signal processing and system theory: theoretical results and some applications*, Proc. IEEE 79(3), 278-305 (1991).
49. A. S. Monin and A. M. Yaglom: *Statistical fluid mechanics* (MIT Press, Cambridge, Mass., 1967), vols. 1 and 2.
50. S. L. Musher, A. M. Rubenchik, and V. E. Zakharov: *Weak Langmuir turbulence*, Phys. Rep. 252, 177-274 (1995).
51. J.-F. Muzy, E. Bacry, and A. Arneodo: *Multifractal formalism for fractal signals: the structure-function approach versus the wavelet-transform modulus-maxima method*, Phys. Rev. E 47, 875-884 (1993).
52. S. W. Nam and E. J. Powers: *Applications of higher order spectral analysis to cubically nonlinear system identification*, IEEE Trans. Acoust. Speech Signal Proc. ASSP-42, 1746-1765 (1994).
53. C. L. Nikias and A. P. Petropulu: *Higher-order spectra analysis* (Prentice Hall, New Jersey, 1993).
54. C. L. Nikias and M. E. Raghuveer: *Bispectrum estimation: a digital signal processing framework*, Proc. IEEE 75, 869-891 (1987).
55. G. Paladin and A. Vulpiani: *Anomalous scaling laws in multifractal objects*, Phys. Reports 4, 147-225 (1987).
56. R. K. Pearson: *Discrete-time dynamic models* (Oxford Univ. Press, Oxford, 1999).
57. H. L. Pécseli and J. Trulsen: *On the interpretation of experimental methods for investigating nonlinear wave phenomena*, Plasma Phys. Controlled Fusion 25, 1701-1715 (1993).
58. D. B. Percival and A. T. Walden: *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Tehcniques* (Cambridge University Press, Cambridge, 1993).
59. P. Pommois, G. Zimbardo, and P. Veltri: *Magnetic field line transport in three dimensional turbulence: Lévy random walk and spectrum models*, Phys. Plasmas 5, 1288-1297 (1998).
60. C. P. Price and D. E. Newman: *Using the R/S statistic to analyze AE data*, J. Atmos. Solar-Terrestr. Phys. 63, 1387-1397 (2001).
61. T. Schreiber: *Interdisciplinary application of nonlinear time series analysis*, report WUB-98-13 (University of Wuppertal, Wuppertal, Germany, 1998) [available at http://xxx.lanl.gov/abs/chao-dyn/9807001].
62. M. F. Shlesinger, G. M. Zaslavsky, and J. Klafter: *Strange kinetics*, Nature 363, 31-37 (1993).
63. T. H. Solomon, E. R. Weeks, and H. Swinney: *Chaotic advection in a two-dimensional flow: Lévy flights and anomalous diffusion*, Physica D 76, 70-84 (1994).
64. L. Sorriso-Valvo, V. Carbone, P. Giuliani, P. Veltri, R. Bruno, E. Martines, and V. Antoni: *Intermittency in plasma turbulence*, Planet. Space Sci., 49, 1193-1200 (2001).
65. K. Stasiewicz, B. Holback, V. Krasnosel'skikh, M. Boehm, R. Boström, and P. M. Kintner: *Parametric instabilities of Langmuir waves observed by Freja*, J. Geoph. Res. 101, 21515-21525 (1996).

66. L. Rezeau, G. Belmont, B. Guéret, and B. Lembège: *Cross-bispectral analysis of the electromagnetic field in a beam-plasma interaction*, J. Geophys. Res. 102, 24387-24392 (1997).
67. C. P. Ritz and E. J. Powers: *Estimation of nonlinear transfer functions for fully developed turbulence*, Physica D 20, 320-334 (1986).
68. C. P. Ritz, E. J. Powers, T. L. Rhodes, et al.: *Advanced plasma fluctuation analysis techniques and their impact on fusion research*, Rev. Sci. Instrum. 59, 1739-1744 (1988).
69. C. P. Ritz, E. J. Powers, and R. D. Bengtson: *Experimental measurement of three-wave coupling and energy cascading*, Phys. Fluids B 1, 153-163 (1989).
70. B. W. Silverman: *Density estimation for statistics and data analysis* (Chapman & Hall, London, 1986).
71. D. Sornette: *Critical phenomena in natural sciences: chaos, fractals, self-organisation and disorder* (Springer Verlag, Berlin 2000).
72. J. Soucek, T. Dudok de Wit, V. Krasnoselskikh, and A. Volokitin: *Statistical analysis of non-linear wave interactions in simulated Langmuir turbulence data*, Ann. Geoph., in press.
73. J. Takalo, R. Lohikoski, and J. Timonen: *Structure function as a tool in AE and Dst time series analysis*, Geophys. Res. Lett. 22, 635-638 (1995).
74. B. P. van Milligen, E. Sánchez, T. Estrada, C. Hidalgo, B. Brañas, B. Carreras, and L. García: *Wavelet bicoherence: a new turbulence analysis tool*, ‘ Phys. Plasmas 2, 3017-3032 (1995).
75. D. V. Vassiliadis, D. N. Baker, H. Lundstedt, and R. C. Davidson (eds.): *Nonlinear methods in space plasma physics*, Phys. Plasmas 6(11), 4137-4199 (1999).
76. D. Vassiliadis: *System identification, modeling, and prediction for space weather environments*, IEEE Trans. Plasma Science, in the press.
77. N. Watkins, M. P. Freeman, S. C. Chapman, and R. O. Dendy: *Testing the SOC hpothesis for the magnetosphere*, J. Atmos. Terr. Phys., 63, 1360-1370 (2001).
78. V. E. Zakharov, S. L. Musher, and A. M. Rubenchik: *Hamiltonian approach to the description of non-linear plasma phenomena*, Phys. Rep. 129, 285-366 (1985).

Subject Index