

Jörg Büchner *Editor*

Space and Astrophysical Plasma Simulation

Methods, Algorithms, and Applications



Space and Astrophysical Plasma Simulation

Jörg Büchner
Editor

Space and Astrophysical Plasma Simulation

Methods, Algorithms, and Applications



Springer

Editor

Jörg Büchner 

Max Planck Institute for Solar System
Research
Göttingen, Germany

ISBN 978-3-031-11869-2

ISBN 978-3-031-11870-8 (eBook)

<https://doi.org/10.1007/978-3-031-11870-8>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

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

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book is dedicated to Maha Ashour-Abdalla, co-founder of the International School and Symposia of Space Simulation (ISSS), who unfortunately left us prematurely. Maha always cared about educating the space simulation abilities of early career stage scientists. She made essential contributions to the formation of a worldwide and now well-established community of highly qualified scientists who apply numerical simulation techniques to better understand space plasmas throughout the Universe.

Göttingen, Berlin and Potsdam in May 2022, Jörg Büchner

Preface

After the 6th International School and Symposium of Space Simulation (ISSS) was held in Garching (Germany), a collection of its lectures was published as Lecture Notes in Physics (Space Plasma Simulation, Springer-Verlag, Berlin, 2003, eds. Büchner, J., Dum, C. T. and Scholer, M.). Since then, the book has been widely distributed and used in many ways used in the space science community as an introduction to numerical simulation methods. Meanwhile space plasma simulation techniques have been successfully utilized to solve open questions concerning the dynamics of the hot space plasmas. As these methods were further developed and extended, they were applied to understanding the most important phenomena of the hot plasmas in the universe. Almost 20 years later the editorial department of Springer's International Publishing felt that a new generation of scientists would greatly benefit from an updated space and astrophysical plasma simulation book, containing actual tutorials of the basic hot plasma simulation approaches, introductions into newly developed methods and algorithms, and examples of modern space and astrophysical applications that became possible with the availability of greatly enhanced computing facilities.

As a result, we, as a group of experts, have combined our accumulated expertise to compile an up-to-date collection of our lectures. Our book aims both at newcomers as well as experienced scientists who want to understand state-of-the-art simulation methods and at those who want to carry out numerical simulations for a better understanding of the complex dynamics of the hot plasmas in the universe. In this sense the chapters of the book can be studied all together as a group but also independently.

Nowadays, numerical simulation is a necessary tool needed to properly understand the highly nonlinear and nonlocal phenomena of collisionless, hot, dilute space and astrophysical plasmas. As a result, in recent years with the development of powerful computers, the numerical simulation of hot plasma phenomena has become indispensable for their thorough investigation. Simulations test theoretical predictions and guide the analysis of large-scale global phenomena, and of in situ observable microscopic processes in space, such as the acceleration of charged particles and properties of whole plasma systems, such as heating of individual plasma components and the emission of energetic electrons, positrons, ions as well as electromagnetic waves over wide ranges of energies and frequencies.

This book introduces in three parts the techniques of the numerical simulation of hot, mostly collisionless, space and astrophysical plasmas. The first part contains tutorials which introduce the most commonly used plasma models, the algorithms of their numerical treatment, and examples of their practical applications. The second part of the book is devoted to advanced simulation techniques which allow more realistic investigations of space and astrophysical hot plasma phenomena. The third part of the book provides two outlooks on algorithms which are presently under development to broaden the scope of applicability of numerical simulation techniques. They are inspired by space plasma simulations but extend far beyond them.

All chapters are written by experts who are internationally well-known for their original research results and contributions to the field. The contents of the book chapters are also based on the year-long experience of the authors teaching these methods at universities in Europe, Japan, and the USA. The whole tutorial course, for example, was taught for many years by members of the Theory and Simulation of Solar System group of the Max Planck Institute for Solar System Research at the Georg-August University in Göttingen and at the Center for Astronomy and Astrophysics of the Berlin Institute of Technology in Berlin, Germany, in order that the feedback of the students could be taken into account.

As the scientific editor, I hereby very much thank all the other authors for their valuable contributions, which helped to make this book so unique. Many thanks also go to Dr. Lisa Scalone of the Springer International Publishing house for her valuable support in technically preparing the volume for print and Dr. Patricio Munoz for his persistent care about the deposited manuscripts in preparation. Because every chapter of the book also underwent a careful peer review process, many thanks also go to the referees, who constructively contributed to improving the content and presentation of the material.

Göttingen, Germany
May 2022

Jörg Büchner

Contents

Part I Introduction to Basic Knowledge—Tutorials

1	Magnetohydrodynamic Simulations	3
	Antonius Otto	
2	Hall Magnetohydrodynamics	29
	J. D. Huba	
3	Hybrid-Kinetic Approach: Massless Electrons	63
	Dan Winske, Homa Karimabadi, Ari Yitzchak Le, Nojan Nick Omidi, Vadim Roytershteyn, and Adam John Stanier	
4	Gyrokinetics	93
	Frank Jenko	
5	Eulerian Approach to Solve the Vlasov Equation and Hybrid-Vlasov Simulations	123
	Francesco Califano and Silvio Sergio Cerri	
6	Fully Kinetic Simulations: Semi-Lagrangian Particle-in-Cell Codes	163
	Patrick Kilian and Felix Spanier	

Part II Introduction to Advanced Simulation Approaches

7	Adaptive Global Magnetohydrodynamic Simulations	211
	Tamas I. Gombosi, Yuxi Chen, Zhenguang Huang, Ward B. Manchester, Igor Sokolov, Gabor Toth, and Bart van der Holst	
8	Multiscale Kinetic Simulations	255
	Giovanni Lapenta	
9	Hybrid-Kinetic Approach: Inertial Electrons	283
	Neeraj Jain, Patricio A. Muñoz, and Jörg Büchner	
10	Generalized Quasi-Neutral Hybrid-Kinetic Simulations	313
	Takanobu Amano	

11 Fully Kinetic (Particle-in-Cell) Simulation of Astrophysical Plasmas	337
Masahiro Hoshino	
Part III Introduction to Advanced New Algorithms and Developments for Future Simulations	
12 Higher-Order Magnetohydrodynamic Simulations	361
Jean-Mathieu Teissier and Wolf-Christian Müller	
13 EMAPS: An Intelligent Agent-Based Technology for Simulation of Multiscale Systems	393
Yuri Omelchenko and Homa Karimabadi	

Contributors

Takanobu Amano University of Tokyo, Tokyo, Japan

Jörg Büchner Max-Planck-Institute for Solar System Research, Göttingen, Germany

Francesco Califano Department of Physics “E. Fermi”, Università di Pisa, Pisa, Italy

Yuxi Chen Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

Silvio Cerri Observatoire de la Côte d’Azur, Boulevard de l’Observatoire, Nice, France

Tamas Gombosi Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

Masahiro Hoshino University of Tokyo, Tokyo, Japan

Zhenguang Huang Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

Joe Huba Syntek Technologies, Arlington County, VA, USA

Neeraj Jain Center for Astronomy and Astrophysics, Technical University Berlin, Berlin, Germany

Frank Jenko Max-Planck-Institute for Plasmaphysics, Garching, Germany

Homa Karimabadi Analytics Ventures, Inc., San Diego, CA, USA

Patrick Kilian Space Science Institute Boulder, CO, USA

Giovanni Lapenta Centre for Plasma Astrophysics, Department of Mathematics, Catholic University of Leuven, Heverlee, Belgium

Ari Yitzchak Le Los Alamos National Laboratory, Los Alamos, NM, USA

Wolf-Christian Müller Center for Astronomy and Astrophysics, Technical University Berlin, Berlin, Germany

Patricio Munoz Max-Planck-Institute for Solar System Research, Göttingen, Germany

Yuri Omelchenko Space Science Institute, Boulder, CO, USA

Nojan (Nick) Omidi Solana Scientific Inc., Solana Beach, CA, USA

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

Vadim Roytershteyn Space Science Institute, Boulder, CO, USA

Felix Spanier Zentrum für Astronomie; Institut für Theoretische Astrophysik, Universität Heidelberg, Heidelberg, Germany

Igor Sokolov Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

Adam John Stanier Los Alamos National Laboratory, Los Alamos, NM, USA

Gabor Toth Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

Jean-Mathieu Teissier Center for Astronomy and Astrophysics, Technical University Berlin, Berlin, Germany

Bart Van der Holst Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

Dan Winske Los Alamos National Laboratory, Los Alamos, NM, USA

Acronyms

ABM	Agent-based model
AEoS	Anisotropic equations of state
AFOSR	Air Force Office of Scientific Research
AFRL	Air Force Research Laboratory
AFWA	Air Force Weather Agency
AGN	Active galactic nucleus
AMR	Adaptive mesh refinement
API	Abstract programming interface
BATS-R-US	Block adaptive tree solar-wind Roe-type upwind scheme
BGK	Bernstein-Greene-Kruskal
CA	Cellular automata
CAM-CL	Current advance method and cycling leapfrog
CCMC	Community Coordinated Modeling Center
CFD	Computational fluid dynamics
CFI	Current filamentation instability
CFL	Courant-Friedrichs-Lowy
CGL	Chew-Goldberger-Low
CHIEF	Code hybrid with inertial electron fluid
CIC	Cloud-in-cell
CISM	Center for Integrated Space Weather Modeling
CPML	Convolutional PML
CR	Cosmic ray
CS	Current sheet
CSEM	Center for Space Environment Modeling
CT	Constrained transport
CWENO	Central weighted non-oscillatory
DES	Discrete-event simulation
DSA	Diffusive shock acceleration
ECPIC	Energy conserving PIC
ECSim	Energy conserving semi-implicit
EFDM	Electron-frame dissipation measure
EMAPS	Event-driven multi-agent planning system
EFDM	Electron-frame dissipation measure
EMHD	Electron magnetohydrodynamics

ENO	Essentially non-oscillatory
EOC	Experimental order of convergence
EP	Energetic particle
EVDF	Electron velocity distribution function
FAC	Field-aligned current
FCB	Foreshock compressional boundary
FCT	Flux-corrected transport
FDTD	Finite-difference time-domain
FFT	Fast Fourier transform
FGMRES	Flexible-generalized minimum residual
FLR	Finite Larmor radius
FMI	Finnish Meteorological Institute
FRB	Fast radio burst
FRC	Field-reversed configuration
FTE	Flux transfer event
GEM	Geospace Environment Modeling
GK	Gyrokinetic
GQNH	Generalized quasi-neutral hybrid
GRB	Gamma-ray burst
GSI	Global smoothness indicators
GSM	General solar magnetic
GUMICS	Grand unified magnetosphere-ionosphere coupling simulation
HFA	Hot flow anomalies
HLL	Harten-Lax-van Leer
HLLD	Harten-Lax-van Leer discontinuity
HVM	Hybrid Vlasov-Maxwell
IMF	Interplanetary magnetic field
IPS	Interplanetary scintillation
ISM	In chapter 7: integrated space-weather model
ISM	In chapter 11: interstellar medium
ISTP	International solar-terrestrial physics
ITG	Ion temperature gradient
LCPFCT	A flux-corrected transport algorithm for solving generalized continuity equations, developed by the NRL
LFM	Lyon-Fedder-Mobarry
LIC	Line integral convolution
LLF	Local Lax-Friedrichs
KAW	Kinetic Alfvén waves
KH	Kelvin-Helmholtz
LIC	Line integral convolution
MAS	Magnetohydrodynamics around a sphere
MHD	Magnetohydrodynamics
MHD-EPIC	MHD embedded PIC
MLMD	Multilevel-multidomain
MMS	Magnetospheric multi-scale (NASA space mission)

MPCDF	Max-Planck-Computing and Data Facility (Germany)
MPI	Message passing interface
MRC	Mission Research Corporation (USA)
MRI	Magneto-rotational instability
NCI	Numerical Cherenkov instability
NGP	Nearest-grid-point
NKZ	Newton-Krylov-Schwarz
NRL	Naval Research Laboratory (USA)
NSF	National Science Foundation (USA)
NSFD	Non-standard finite-difference
PDE	Partial differential equation
PDM	Partial donor method
PEC	Perfect electric conductor
PEP	Preemptive event processing
PIC	Particle-in-cell
PML	Perfectly matched layer
PPC	Particles per cell
PQS	Piecewise quadratic shape
PSATD	Pseudo-spectral analytic time domain
PSD	Phase space density
QLT	Quasi-linear theory
QNTF	Quasi-neutral two-fluid
RoR	Runs-on-request
SIMD	Single-instruction-multiple-data
SMI	Synchrotron maser instability
SNR	Supernova remnant
SSPRK	Strong stability preserving Runge-Kutta
STOS	Simulation time operating system
SWMF	Space weather modeling framework
SWPC	Space Weather Prediction Center
TAE	Toroidal Alfvén eigenmodes
TD	Tangential discontinuity
TSC	Triangular shaped cloud
TVD	Total variation diminishing
TVD-RK	Total variation diminishing Runge-Kutta
UAV	Unmanned aerial vehicles
UCT	Upwind constrained transport
UHECR	Ultra-high-energy cosmic ray
VAC	Versatile advection code
WDC	World Data Center
WENO	Weighted, essentially non-oscillatory
WFA	Wakefield acceleration
WI	Weibel instability
WSA	Wang-Sheeley-Arge

Part I

Introduction to Basic Knowledge—Tutorials

Six tutorials introduce the most commonly used models used to describe the mostly hot, practically collisionless space and astrophysical plasmas, algorithms of their numerical treatment, as well as some typical applications. Chapter 1 is devoted to the magnetohydrodynamic (MHD-) approach to the description of plasmas as a fluid. The MHD model obeys the laws of mass, momentum, and energy conservation, but neglects the details of the plasma microphysics, the motion and energization of particles, kinetic plasma instabilities, and strong anisotropies, and the causes of energy dissipation. So they are most appropriate to describe large- and meso-scale phenomena, such as fluid instabilities and magnetic reconnection. However, for the understanding of the underlying microphysics, more detailed plasma models have to be employed. Hence, the remaining five tutorial chapters consider increasingly refined plasma models that take into account more and more aspects of the plasma dynamics. The often important role of electrons that are much lighter than the ions can be described in its simplest form by means of the so-called Hall-magnetohydrodynamical approximation (Chap. 2). If the properties of ions as particles are essential, while the mass of the electrons can be neglected, hybrid-kinetic simulations with a massless electron fluid can be carried out, as presented in Chap. 3, tracing ions as macro-particles using a particle-in-cell (PIC) approach or, as discussed in Chap. 5, by directly solving the Vlasov equations for the ions by an Eulerian approach. The most complete fully kinetic description of hot plasmas, however, needs to solve the Vlasov equations for both ions and the much lighter electrons. This has to be done including all six phase space dimensions while well resolving the velocity space, either by an Eulerian solution of the Vlasov equation or following a large number of macro-particles by a semi-Lagrangian PIC method. Needless to say, such simulations are very expansive. However, in the presence of strong magnetic fields, the phase space dimension can be reduced by one, by averaging over the gyrophase of the particles. The corresponding gyrokinetic approach, which originally was developed for magnetically confined fusion plasmas, is introduced in Chap. 4. Chapter 5 introduces a way to directly solve the kinetic (Vlasov-) equations by an Eulerian approach, allowing, for example, hybrid-Vlasov code simulations of collisionless plasmas. The tutorial part of this book is completed by Chap. 6, an introduction into fully kinetic plasma simulations that utilize semi-Lagrangian particle-in-cell (PIC-) codes.



Magnetohydrodynamic Simulations

1

Antonius Otto

Abstract

MHD (magnetohydrodynamic) simulations have been used to model mesoscale and global plasma dynamics for more than 40 years. Applications range from laboratory plasmas and magnetospheric and solar dynamics to astrophysical systems such as interstellar ionized gases. This success is due to the self-consistent local conservation of mass, momentum, and energy—a basic requirement to model macroscale plasma systems. MHD modeling avoids the high complexity of microscale physics, and with powerful computational environments, it has become the basic tool to examine plasmas. MHD is not applicable if microscale aspects such as charged particle motion, kinetic instabilities, or strong anisotropies play an important role. This chapter provides a discussion of basic concepts and initial and boundary conditions for MHD simulations. These are illustrated with three specific examples: (1) Kelvin-Helmholtz modes at magnetospheric boundaries, (2) the interaction of fast (bow) shocks with upstream low-density perturbations which create strong transient events upstream of the original shock, and (3) the slow evolution of magnetotails that leads to the formation of thin current sheets before an eruptive energy release. A summary and discussion provide further guidance into the use of MHD simulations.

A. Otto (✉)

University of Alaska Fairbanks, Geophysical Institute, Fairbanks, AK, USA
e-mail: otto3@alaska.edu

1.1 MHD Equations and Properties

Most of the visible universe is in the plasma state where the kinetic energy of charged particles is much larger than their potential energy. Examples in our own solar system are the atmosphere of our sun and the stream of charged particles from the sun called solar wind which interacts with unmagnetized and magnetized bodies in the solar system. In the latter case, this creates planetary magnetospheres of which our own terrestrial magnetosphere is the best-studied example. At large distances from the sun, the solar wind interacts with the interstellar neutral and charged material.

In general, properties and evolution of a plasma involve electromagnetic interaction and are much more complicated than that of a hydrodynamic system. Particularly on small spatial scales, aspects of charge neutrality and gyromotion of charged particles complicate the plasma dynamics. Fortunately, many systems such as solar magnetic structures and most planetary magnetospheres are much larger than typical kinetic scales. For instance, the terrestrial magnetosphere has typical scales of 10s of Earth radii ($1R_E \approx 6400\text{ km}$), while ion gyroscales are a few 100 km, and electron scales are much smaller. This argument does not imply that kinetic physics can be ignored, but rather that errors due to kinetic physics on the much larger magnetospheric scales are often small. Therefore, many large-scale plasma systems can be understood in terms of a collective fluid plasma description called magnetohydrodynamics. The formulation of the magnetohydrodynamic (MHD) equations by Alfvén, Schlueter and others enabled the theoretical understanding of many plasma systems such as magnetospheres, solar atmospheric structures, and laboratory applications from the 1950s.

Plasma fluid equations can be systematically derived by integrating the moments of the Boltzmann equation in velocity space which yields equations for mass, momentum, and energy density in space for electrons and each ion species. Assuming a single ion species only and charge neutrality and defining a total mass density ρ and average bulk velocity \mathbf{u} through $\rho = n(m_i + m_e)$ and $\rho\mathbf{u} = n(m_i\mathbf{u}_i + m_e\mathbf{u}_e)$, where m_e and m_i are electron and ion mass, n is number density, and \mathbf{u}_e and \mathbf{u}_i are electron and ion bulk velocity, the plasma fluid continuity and momentum equations become

$$\frac{\partial\rho}{\partial t} + \nabla \cdot \rho\mathbf{u} = 0 \quad (1.1)$$

$$\frac{\partial\rho\mathbf{u}}{\partial t} + \nabla \cdot (\rho\mathbf{u}\mathbf{u}) + \frac{m_e m_i}{e^2} \nabla \cdot \left(\frac{1}{\rho} \mathbf{j} \mathbf{j} \right) = -\nabla p + \mathbf{j} \times \mathbf{B} \quad (1.2)$$

$$\mathbf{E} + \mathbf{u} \times \mathbf{B} - \frac{m_e m_i}{e^2 \rho} \left[\frac{\partial \mathbf{j}}{\partial t} + \nabla \cdot (\mathbf{u} \mathbf{j} + \mathbf{j} \mathbf{u}) \right] = -\frac{M}{e \rho} \nabla p_e + \frac{m_i}{e \rho} \mathbf{j} \times \mathbf{B} + \eta \mathbf{j} \quad (1.3)$$

(e.g., [1]). Here, e represents the elementary charge, \mathbf{j} current density, p the sum of electron and ion pressure, \mathbf{B} the magnetic field, \mathbf{E} the electric field, $M = m_e + m_i$, and η resistivity. The first equation represents the continuity equation for mass,

the second is the equation for the bulk momentum, and the third is an equation for the current density or electron momentum better known as generalized Ohm's law. Applying scaling laws in Eq. (1.3), the term in square brackets scales with the electron inertial scale λ_e^2/L_0^2 where $\lambda_e = c/\omega_{pe}$ represents the electron inertial scale, ω_{pe} the electron plasma frequency, and L_0 a typical scale length of the plasma system under consideration. The first two terms on the right side of generalized Ohm's law (3) scale with the ion inertial scale λ_i/L_0 with $\lambda_i = c/\omega_{pi}$ where the $\mathbf{j} \times \mathbf{B}$ term is addressed as the Hall term. For the solar corona, λ_e is a few cm or less, and for the magnetosphere, λ_e is of the order of several 100 m to a few km such that for almost all applications the corresponding term can be ignored in generalized Ohm's law. This is the case also for the \mathbf{jj} term in the momentum equation which scales with the electron inertial length and provides the electron contribution to the bulk momentum equation. Although the ion inertia scale is much larger than the corresponding electron scale $\lambda_i^2/\lambda_e^2 = m_i/m_e$, it is often still much smaller than typical scales of the system, particularly for solar applications. Note also that the electron velocity can be expressed as $\mathbf{u}_e = \mathbf{u} - \mathbf{j}/ne$ where the current density term scales with λ_i/L_0 such that the electron velocity can be significantly different from the ion of bulk plasma velocity when the ion inertial scale is important (Hall physics).

Ignoring the electron and ion inertial terms on sufficiently large scales, Eq. (1.3) becomes the well-known resistive Ohm's law. The three equations are complemented by equations for energy density of electrons and ions and Maxwell's equations for a neutral medium. Strictly the pressure terms in Eqs. (1.2) and (1.3) should be the divergence of the corresponding pressure tensors. However, to lowest order, these pressure terms just represent the trace of the respective pressure tensor or thermal energy. With plasma isotropy and neglecting kinetic and inertial plasma scales, the fluid equations are the so-called MHD equations:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \mathbf{u} = 0 \quad (1.4)$$

$$\frac{\partial \rho \mathbf{u}}{\partial t} + \nabla \cdot (\rho \mathbf{u} \mathbf{u}) = -\nabla p + \mathbf{j} \times \mathbf{B} \quad (1.5)$$

$$\mathbf{E} + \mathbf{u} \times \mathbf{B} = \eta \mathbf{j} \quad (1.6)$$

$$\frac{1}{\gamma - 1} \left(\frac{\partial}{\partial t} p + \nabla \cdot p \mathbf{u} \right) = -p \nabla \cdot \mathbf{u} + \eta \mathbf{j}^2 \quad (1.7)$$

(e.g., [2,3]). Sometimes, the last equation for pressure or thermal energy is replaced by an equation for total energy density for strict local energy conservation. The assumption of isotropy strictly requires either a sufficient number of collisions or an unmagnetized plasma. However, for many macroscale problems, local mass, momentum, and energy conservation seem more important on sufficiently large

scales than plasma isotropy. The equations are complemented by Ampere's law and the induction equation:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} \quad (1.8)$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \quad (1.9)$$

where the displacement current is ignored for nonrelativistic applications. Here, $\nabla \cdot \mathbf{B} = 0$ is introduced through the initial conditions, and the induction equation insures that it should be satisfied for all times.

It is worth noting that ideal Ohm's law $\mathbf{E} + \mathbf{u} \times \mathbf{B} = 0$ implies the so-called frozen-in condition, i.e., for any closed contour which moves with the bulk velocity \mathbf{u} , the amount of magnetic flux through this contour does not change with time (e.g., [4]). This condition can also be expressed in a different manner, specifically, if two fluid elements that move with the velocity \mathbf{u} are connected through a magnetic field line at one time they are connected through a field line at all times as illustrated in Fig. 1.1. Note that on sufficiently small scales with the Hall term included, ideal Ohm's law is $\mathbf{E} + \mathbf{u}_e \times \mathbf{B} = 0$ such that the frozen-in condition applies to the electron fluid velocity \mathbf{u}_e instead of the bulk velocity \mathbf{u} .

In addition to this concept of magnetic flux conservation, there are a number of basic conserved quantities which are mass, momentum, and energy conservation. These conservation laws can be expressed in a straightforward manner where the spatial derivatives are expressed as gradients or the divergence of tensor quantities:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \rho \mathbf{u} \quad (1.10)$$

$$\frac{\partial \rho \mathbf{u}}{\partial t} = -\nabla \cdot \left[\rho \mathbf{u} \mathbf{u} + \left(p + \frac{B^2}{2\mu_0} \right) \mathbf{I} - \frac{1}{\mu_0} \mathbf{B} \mathbf{B} \right] \quad (1.11)$$

$$\frac{\partial w_{tot}}{\partial t} = -\nabla \cdot \left[\left(\frac{1}{2} \rho u^2 + \frac{\gamma p}{\gamma - 1} + \frac{1}{\mu_0} B^2 \right) \mathbf{u} - \frac{\mathbf{u} \cdot \mathbf{B}}{\mu_0} \mathbf{B} + \frac{\eta}{\mu_0} \mathbf{j} \times \mathbf{B} \right] \quad (1.12)$$

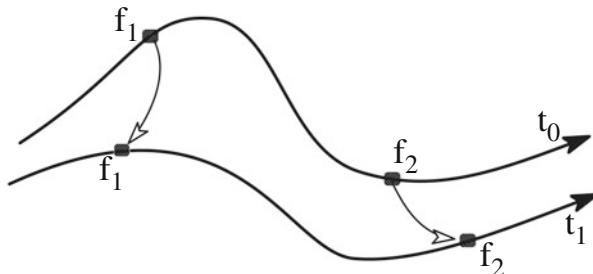


Fig. 1.1 Illustration of magnetic line conservation for fluid elements f_1 and f_2

with

$$w_{tot} = \frac{1}{2}\rho u^2 + \frac{p}{\gamma - 1} + \frac{1}{2\mu_0}B^2$$

For instance, mass for any fixed finite volume V_0 where with mass $M = \int_{V_0} \rho d^3r$ is then expressed as $dM/dt = \oint_{\delta V} \rho \mathbf{u} \cdot d\sigma$ where $d\sigma$ is a surface element of the boundary δV of the volume V . This illustrates that changes in density can occur only with momentum density flux through the surface of the chosen volume and require a component of the velocity normal to that surface. This also implies that mass is conserved in any volume moving with the fluid. Similar conclusions subject also to the magnetic field orientation at the boundary can be reached for momentum and energy conservation.

Note that the substitution $p = h^\gamma$ in the pressure equation (7) generates a continuity equation which is exactly satisfied for ideal conditions ($\eta = 0$):

$$\frac{\partial h}{\partial t} + \nabla \cdot h \mathbf{u} = \frac{\gamma - 1}{\gamma} h^{1-\gamma} \eta \mathbf{j}^2 \quad (1.13)$$

However, this formulation requires a continuous and differentiable value of the pressure and is violated for the case of shocks and MHD discontinuities. It still provides an interesting test as an ideal conservation law.

Conservation laws and particularly the frozen-in condition for ideal MHD are of major importance for any plasma simulation. Most space plasma systems are collisionless such that collisional resistivity is really 0. However, at very thin boundaries with strong currents, collective processes can exchange ion and electron momentum yielding an effective resistivity or some other terms in generalized Ohm's law which violates the frozen-in condition. It should be emphasized that this violation occurs in a very small and limited volume with extreme gradients, while almost everywhere the frozen-in condition is satisfied. This is important because, in some way, many space plasma problems address the transport of magnetic flux or mass, momentum, and energy across boundaries. This transport requires the local violation of the frozen-in condition. Therefore, any significant violation due to numerical errors can easily invalidate the results from numerical simulations (not only in MHD systems).

1.2 Basic Considerations for the Numerical Solution of MHD Systems

The set of MHD equations (1.4)–(1.9) represents a set of coupled partial differential equations (PDEs). A few special problems may have analytic solutions, but most other applications require a numerical treatment including a discretization of these equations, which typically generates a set of algebraic equations, and requires a method to solve the discretized equations (e.g., [5, 6]). It is obvious that the

algebraic equations are not the same as the original PDEs and therefore their solution must be expected to be different. However, a necessary condition for any chosen discretization must be consistent with the PDEs with increasing accuracy of the discretization. In other words, in the limit where grid element sizes and time step size goes to 0, the original PDEs must be recovered from the algebraic equations. Otherwise, it cannot be expected that the numerical solution converges to the exact solution of the PDEs.

There is a large range of different discretization methods such as finite difference, finite volume, finite element, or spectral methods, all of which attempting to minimize errors from the spatial discretization or decomposition into base functions. Specifically, spectral methods use certain (appropriate) classes of base functions to discretize PDEs. Finite element methods decompose the simulation domain into many smaller, geometrically adequate subdomains and use base functions in these subdomains to represent the PDEs. Similarly, finite volume methods use many smaller sub-volumes to represent derivatives and fluxes associated with the PDEs in these sub-volumes to minimize numerical errors in the sub-volumes. Finite difference methods use many grid points in a structured or unstructured grid to represent spatial derivatives in PDEs. Note that each of these different approaches has advantages and disadvantages that depend on the choice of the application, geometry, etc. For instance, particularly the finite element and finite volume methods have advantages when the physics involves a particularly complicated shape of the simulations domain, for instance, flow around a specific obstacle and associated thin boundary layer regions. However, because of the limited scope and because of their frequent use, this article will focus mostly on finite difference methods.

In general, the numerical solution of PDEs depends very much on the type of the PDEs. Sufficiently simple systems of PDEs can be classified into three groups which are elliptic, parabolic, and hyperbolic (e.g., [6]). A typical example for elliptic PDEs is the Poisson equation, and applications that generate elliptic PDEs are, for instance, equilibrium problems or supersonic flow. There is no temporal evolution, and a small change of the boundary conditions or of the sources has an influence on the solution in the entire domain. The solution is often found through iterative methods and involves the inversion of a large matrix representing the system of equations. An example for parabolic equations is the diffusion equation, and applications usually address problems where diffusion and viscosity are dominant for a problem. Again, changes in a small domain at the boundary or in the diffusion coefficients have an instant influence on the entire system. Most common, particularly for the MHD equations, are hyperbolic PDEs. These equations are typically time-dependent, involve the transport of information with a finite velocity, and have a number of different wave modes, such as sound waves in the case of hydrodynamics, that propagate along the so-called characteristic surfaces, which for linear problems can be used to construct the exact solution for proper initial and boundary conditions.

Ideal MHD has three different basic wave modes (fast, slow, and Alfvén waves) along which information (and momentum and energy) can be transported. Note also

that Eq. (1.13) in combination with the continuity equation can be used to show that the quantity $s = p/\rho^\gamma$ can be expressed as $\partial s/\partial t + \mathbf{u} \cdot \nabla s = 0$ in ideal MHD where s is regarded as an entropy measure. Therefore, this entropy measure is conserved in the frame advecting with the fluid which is sometimes addressed as an entropy wave. Similarly, transport associated with magnetic flux conservation as discussed in relation to Fig. 1.1 can be regarded as a “magnetic flux mode.” However, it should be emphasized that the basic classification into three types is an oversimplification. For instance, the MHD equations can contain aspects of elliptic or parabolic PDEs although they are mostly and for most applications hyperbolic.

Particularly for hyperbolic and parabolic PDEs but also for iterative solutions of elliptic PDEs, a critical property for the solution is stability (e.g., [5,6]). To illustrate this concept, consider a simple transport or advection equation $\partial f/\partial t + u \partial f/\partial x = 0$ with a prescribed constant value of the velocity u . Consider a simple finite difference scheme $f_j^{n+1} = f_j^{n-1} - u \Delta t / \Delta x (f_{j+1}^n - f_{j-1}^n)$ where the upper index n describes a time index with time levels separated by Δt and the lower index is a grid index for the x direction with Δx as the separation of grid points. In other words, the temporal update uses the time step and the spatial finite difference between points adjacent to grid point j . Any perturbation of f propagates with velocity u and takes the time $\tau_{grid} = \Delta x/u$ to travel a single grid spacing. Therefore, if $\Delta t > \tau_{grid}$, information at grid index j must have arrived from beyond the adjacent grid points which, however, is not considered in the equation for the update (using only next neighbor information). Therefore, a time step $\Delta t > \tau_{grid}$ must generate a meaningless and erroneous result which is reflected in this scheme becoming unstable for $\Delta t > O(\tau_{grid})$ leading to exponential growth of any perturbation. This concept is formalized in a stability analysis where the growth or damping of a perturbation is analyzed for the chosen numerical method.

Although there is no general proof that consistency of the discretization and stability of the method insure that the numerical solution converges to the exact solution of the system of PDEs, they are strong arguments for a sound methodology. Also, convergence of the numerical solution to a single solution for successive refinements of the grid and time steps can be used to demonstrate convergence to the actual solution. There are a few more pitfalls that one should be aware of for MHD simulations: Time independency and diffusion/viscosity are parts of MHD so the methodology must be appropriate to solve not only the hyperbolic aspects of the MHD equations part but also the steady-state and dissipative aspects of the MHD equations in a stable manner. Here, complications for nonlinear steady-state problems include multiple solutions for given boundary conditions or for slowly changing boundary conditions. Also, solutions may approach bifurcation points where they can cease to exist and/or develop singularities. Furthermore, even small modifications such as the inclusion of the Hall terms can require very small integration steps, in this case because the Hall term enables the fast propagating whistler wave. Similarly, modifications of the basic equations that include elliptic terms can require very small iteration steps to generate a solution, a property that is addressed as stiffness.

Additional insight on properties of a discretization can be obtained through an expansion of the discretized algebraic equations which in the case of the advection equation assume a general form $\partial f / \partial t + u \partial f / \partial x = \sum_{i=1} c_i \Delta x^{i-1} \partial^i f / \partial x^i$. Clearly, consistency requires $c_1 = 0$; otherwise, the equation is altered from the original PDE. However, the expansion demonstrates the presence of second- and higher-order even derivatives that act as numerical dissipation (diffusion). Similarly, odd order derivatives cause numerical dispersion. Both effects are largest on the grid scale and are more strongly localized for higher-order accurate schemes, i.e., where $c_i = 0$ for a range of low values of i . In general, more accurate discretizations do not eliminate issues with numerical diffusion and dispersion. Although numerical diffusion is bad because it diffuses structures and widens plasma boundaries, numerical diffusion is unavoidable and acceptable if it is sufficiently localized. In fact, numerical diffusion is required to balance effects of numerical dispersion. Numerical dispersion of sharp plasma boundaries leads to large and growing oscillations on the grid scale. Such oscillations can easily cause the termination of any simulation because of locally negative density or pressure. Such grid oscillations are avoided if numerical diffusion is present to suppress them. Therefore, numerical diffusion should be a lower-order error than the dispersion. This principle is actually employed in the so-called flux-corrected transport (FCT) schemes which attempt to avoid issues that can occur with the presence or evolution of shocks. Since shocks develop as real discontinuities in ideal MHD, they can never be resolved with a finite grid. Therefore, FCT schemes typically use a multistep discretization where a low-order finite difference approach is combined with a high-order scheme. The low-order discretization is retained and applied in locations where the high-order scheme generates large dispersion errors due to discontinuous plasma conditions (e.g., [6]).

There are many tests that can be used to examine any particular MHD simulation scheme. Clearly, wave propagation of different MHD waves (slow, fast, and Alfvén waves) provides an important test. A similarly interesting test is the simulation of the corresponding MHD discontinuities and shocks. Particularly fast shocks are important for the interaction of the superfast solar wind plasma with the magnetosphere. Slow shocks and rotational discontinuities are of importance for problems of magnetic reconnection where much of the plasma acceleration but also dissipation can occur at slow shock structures.

In order to examine more general properties of simulations concerning nonlinear MHD physics, a number of macroscopic instabilities are sufficiently well understood to serve as test cases for MHD simulations. Specifically, instabilities based on strong velocity shear (Kelvin-Helmholtz modes), magnetic shear (tearing modes and reconnection), and gravity-driven plasma interchange (Rayleigh-Taylor modes) are quite well understood (e.g., [3, 7, 8]) and can be used to examine properties and behavior of MHD simulations.

Particularly for problems involving magnetic flux, mass, momentum, and energy transport, tests using the general conservation properties of MHD equations can be instructive. For instance, tests may evaluate the change of total mass, momentum, and energy considering the actual flux at the boundary of the respective domain.

For ideal MHD, magnetic line conservation (or the frozen-in condition) can be examined by using tracers that move with the simulated fluid velocity and that originally started on the same field line. Their separation after any finite time and travel distance can provide a measure of how well fluid elements are frozen to a particular magnetic field line. Magnetic flux tube properties can also be used to examine mass conservation in a moving geometry. Since magnetic field lines move by necessity with the fluid in ideal MHD, the mass on a flux tube with a differential cross section can alter only due to inflow or outflow at the ends of a field line. The mass of a differential flux tube is expressed as the integral $M_B = \int_B n ds / B$ along a magnetic field line between fixed points a and b where n is number or mass density, ds is the line element along the field line, and B is the magnitude of the local magnetic field. By considering inflow and outflow, we can determine how well flux tube mass is conserved. With these definitions, we can address the quantity $V_B = \int_B ds / B$ as the differential flux tube volume of a field line between two points a and b along this field line.

Conservation of M_B can be demonstrated using the continuity equation for n . By the same notion, it can be demonstrated that $S_B = \int_B (p^{1/\gamma}) ds / B$ is conserved in the absence of inflow or outflow at the ends of a field lines (for ideal MHD). If velocities are small such that the system is in an approximate equilibrium, pressure p is approximately constant along the field line such that $S_B = p^{1/\gamma} V$ can be considered as a flux tube entropy measure which is conserved in ideal MHD. As mentioned before, a local measure of this quantity is $s \equiv p^{1/\gamma} / n$ which is constant along the path of a fluid element in ideal MHD and provides an easy test for mass and entropy conservation in MHD simulations. Obviously, all of these conservation laws are not exactly satisfied in numerical simulations. However, they have to be satisfied sufficiently well on the time scale of a simulation such that a numerical solution can distinguish between an actual transport through reconnection or diffusion and a numerical violation of conservation laws.

Finally, it is important to understand that the ideal MHD equations have no intrinsic scale. They are valid from galactic down to small laboratory scales such that the actual evolution of a system is scale independent. Therefore, the MHD equations are often solved in a normalized form, where magnetic field, density, and length scales are measured in typical values, i.e., B_0 , n_0 , and L_0 . With this choice, the natural scale for velocity and pressure becomes $u_0 = B_0 / \sqrt{\mu_0 m_0 n_0}$ and $p_0 = B_0^2 / \mu_0$.

1.3 Initial and Boundary Conditions

Boundary conditions for MHD simulations are notoriously difficult. Boundary conditions depend on the particular application and initial conditions. Particularly a boundary with nonuniform plasma conditions can be challenging. Also, the question of whether there is a plasma inflow or outflow plays a major role. Sometimes, it is claimed that a boundary uses the so-called open boundary conditions, implying that

it allows inflow or outflow just as the system demands. It should be noted that in case of nonuniform boundaries, there is no simple open boundary condition. Strictly, plasma outflow and inflow that are caused by dynamics within the simulation domain can be computed through the characteristics of the system. They determine where and how perturbations from within the simulation demand inflow or outflow at the boundaries. However, while this can be done in principle, MHD has three different anisotropic wave modes, such that this approach is extremely difficult in two- or three-dimensional MHD. Therefore, such boundary conditions often represent a guess that, when it works, is likely the best one can achieve for a boundary condition. Sometimes, it may be necessary to choose boundary locations far from the actual physical process to insure that the boundary has no influence on the evolution of this process.

In the following, we will introduce and discuss boundary and initial conditions for several applications of numerical simulations presented in Sect. 1.4. Specifically, we consider three different applications. An example of a relatively simple and periodic instability is the Kelvin-Helmholtz (KH) instability provided in Sect. 1.4.1. Shock physics is addressed in Sect. 1.4.2 with an example that applies to hot flow anomalies (HFAs) at planetary bow shocks. A more involved boundary is introduced in Sect. 1.4.3 presenting a magnetotail simulation of current sheet thinning. The corresponding initial and boundary conditions are summarized in the next subsections.

1.3.1 Kelvin-Helmholtz Instability

The first magnetopause application is the Kelvin-Helmholtz wave, a straightforward and well-understood instability driven by sufficiently fast shear flow across a plasma boundary. The instability has been studied extensively in two and three dimensions (e.g., [8–12]). This instability operates in two dimensions, and our geometry uses the x direction normal to the boundary and y along the boundary. The initial conditions use for the velocity:

$$u_y(x) = u_0 \tanh x/L_0$$

and $u_x = u_z = 0$; for density $\rho = \rho_0$; for the magnetic field components $b_x = 0$, $b_y = b_{y0}$, and $b_z = b_{z0}$; and for pressure $p = p_0$, where all quantities with an index 0 are assumed constant. Note that instead of a constant magnetic field along y and z , we could have chosen some x -dependent profiles with the only condition that pressure has to be chosen such that $p + b_y^2 + b_z^2 = \text{const}$ such that the configuration is in a steady (time independent) state. For the results in Sect. 1.4, we have chosen $\rho_0 = 1$, $p_0 = 1$, $u_0 = 1$, $b_{y0} = 0.1$, $b_{z0} = 1$, and $L_0 = 1$. None of these values is critical except that the Alfvén speed associated with b_{y0} must be smaller than u_0 for the instability to operate.

The most reliable and straightforward boundary conditions for MHD applications are given by symmetry conditions or configurations that are periodic in space. Sim-

ple applications where periodic conditions can be applied are, for instance, macro-instabilities, such as magnetic tearing modes, or flow-driven Kelvin-Helmholtz modes. Therefore, the simulation here uses periodic conditions along the y direction.

The method used for the applications here is second-order accurate and uses three grid points to evaluate first- and second-order derivatives. Therefore, our simulation grid only requires one layer of guard cells around the physical cell boundaries to apply von Neumann derivative boundary conditions, such that $j = 2$ and $j = n_y - 1$ correspond to physical boundary cells and $j = 1$ and $j = n_y$ to the mathematic boundary cell to apply boundary conditions. Thus periodicity at the y minimum boundary implies $f_{j=1} = f_{j=n_y-2}$ and $f_{j=2} = f_{j=n_y-1}$ for all x indices i .

Boundary conditions at x_{min} and x_{max} use all 0 derivatives normal to the boundary, which explicitly implies, for instance, at the x_{min} boundary $f_{i=1} = f_{i=3}$ for all quantities with one exception. The normal magnetic field derivative that determines $b_{x,i=1}$ and $b_{x,i=n_x}$ computes these values using the tangential field and $\nabla \cdot \mathbf{b} = 0$. Note that $\nabla \cdot \mathbf{b} = 0$ is required for the initial configuration and should be conserved through the integration method. However, it still leaves the possibility that a nonzero divergence is carried into the system through boundaries. This is avoided by employing $\nabla \cdot \mathbf{b} = 0$ at the boundaries where possible.

A zero normal derivative boundary condition is used for the tangential velocity, density, and pressure because they are nonzero in the initial condition at the x boundaries (and almost constant). The normal velocity component can use the same conditions or make use of a $\nabla \cdot \mathbf{v} = 0$. Finally, we note that the conditions at the x boundaries is not an exact symmetry situation and it is difficult to formulate a condition that represents truly free flow for this boundary. Therefore, the actual location of the x boundary is chosen at a large distance from the location of the instability such that the detailed boundary conditions do not have a significant influence on the instability. This can be numerically slightly more expensive, but the applied scheme uses a nonuniform grid with larger grid spacing away from the $x = 0$ plane where the KH instability operates.

1.3.2 Fast Shock Simulation

This second application examines the interaction of a fast shock with a magnetic flux tube with a lower density than the surrounding upstream plasma. The physical motivation for this setup is the occasional observation of HFAs upstream of the Earth's bow shock, during which the plasma is strongly (and nonadiabatically) heated and strongly decelerated, sometimes so much that it is actually streaming sunward. Observations of the so-called density cavities show that the upstream plasma contains sometimes flux tubes with much lower density than the ambient material. Such flux tubes must cause large perturbations at the bow shock of the Earth. These simulations are also two-dimensional. Here, x is the direction normal to the original fast shock into the downstream region, and y is the transverse direction. Fast shock simulations with application to HFAs have been conducted with hybrid simulations (e.g., [13–15]).

The initial state is the superposition of two configurations. First, the simulation uses the solution to the fast shock jump conditions as given by:

$$\begin{aligned} & \left(u_u^2 - X u_A^2 \right)^2 \left[X c_s^2 + \frac{1}{2} u_u^2 \cos^2 \theta \{ X (\gamma - 1) - (\gamma + 1) \} \right] \\ & + \frac{1}{2} u_A^2 u_u^2 X \sin^2 \theta \left[(\gamma + X (2 - \gamma)) u_u^2 - X u_A^2 ((\gamma + 1) - X (\gamma - 1)) \right] = 0 \end{aligned}$$

(e.g., [4]) where $X = \rho_d / \rho_u$ is the shock compression, u_u is the upstream velocity, u_A and c_s are the upstream Alfvén and Sound speed, θ is the angle of the upstream magnetic field with the shock normal, and $\gamma = 5/3$ is the ratio of specific heats. The resulting initial condition can be expressed as:

$$f = f_u + 0.5 \Delta_f [1 + \tanh(x/\delta)] \quad (1.14)$$

with f representing density ρ , velocity components u_x and u_y , magnetic field component b_y , and pressure p . Here, $\Delta_f = f_d - f_u$ represents the jump from the down- to the upstream value, for instance, $\Delta_\rho = \rho_d - \rho_u = \rho_u (X - 1)$. Note this choice centers the fast shock at $x = 0$ and the width of the step function δ adjusts itself immediately after turning on the simulation because it really is determined by the dissipation either of a physical resistivity and viscosity or a corresponding numerical dissipation. In order to examine the interaction of a density depleted flux tube with such a shock, we superpose a density depletion with a shape of $\rho_\delta = 1 - 0.5 \rho_{\delta 0} [\tanh((y_2 - y)/\delta_y) + \tanh((y - y_1)/\delta_y)]$. For $\delta_y > 0$ and $y_2 - y_1 > 2\delta_y$, this function is 1 everywhere except between y_1 and y_2 where it assumes a value of $1 - \rho_\delta$. So by multiplying the initial condition (13) with ρ_δ in the upstream region, we can generate a reduced density region between y_1 and y_2 . Note that y_1 and y_2 are dependent on the x coordinate because we need the depleted density region within a flux tube which is inclined with the boundary normal by an angle θ_{bn} . Note that basic free parameters for this type of study are the shock Mach number, the plasma β , and the angle of the incident magnetic field with the shock normal. Other parameters that can be adjusted are the relative magnitude of the density depletion and the width of the density depleted flux tube.

Initial conditions for this case are relatively straightforward for the boundaries in x . Since x points into the downstream direction, the boundary conditions at the x_{min} boundary are given by Dirichlet conditions where all boundary values are set to the respective upstream region values defined in the initial conditions. One exception is density where the boundary values are computed from the superposition of upstream density and depleted flux tube values. Note that in the normal incident frame where u_y in the upstream region is 0, a flux tube marked by a lower density moves along the y direction when observed at a fixed x coordinate. Therefore, the depleted density region must move along y at the x_{min} boundary with a speed that depends on the upstream velocity and the angle θ_{bn} of the magnetic field. Boundary conditions in the downstream region can be chosen in a similar manner, i.e., as

Dirichlet conditions based on the analytic values from the shock jump conditions. There will always be some reflection because one has to expect at least small deviations of the numerical shock solution from the exact analytic values. However, worse are the conditions caused by the interaction of the low-density flux tube with the shock. These represent large perturbations relative to the original Dirichlet conditions. Therefore, the downstream system size is chosen large enough that these perturbations cannot travel back to the shock region on the considered time scales. Since truly free boundary conditions are very difficult to achieve, the boundary conditions in the y direction are assumed periodic. The density-depleted flux tube size is small compared to the system size in y , such that periodic conditions in y should be close to open boundary conditions.

1.3.3 Magnetotail Simulation

This is an example of a three-dimensional simulation of a part of the Earth's magnetotail. The physical motivation for this simulation is the examination of the formation of very thin current sheets in the near Earth tail region. Such current sheets can be associated with the onset of major instabilities that release large amounts of stored magnetic energy and cause geomagnetic substorms and the associated aurora. Different from the prior two examples, this simulation does not strictly use an analytic initial condition. Rather, it uses an initial condition that is based on a typical magnetic field of the magnetotail. This magnetic field is used with a numerical relaxation to obtain force balance and to generate a three-dimensional magnetotail equilibrium. Coordinates used for this simulation are so called GSM (general solar magnetic) coordinates with x pointing sunward, z northward, and y along the cross-tail direction.

It has been hypothesized that different from a compression of the magnetotail lobes, also the removal of closed magnetic flux in the near Earth tail can lead to the formation of strong current sheets. In order to test and demonstrate this hypothesis, it is required to use a sufficiently realistic magnetic field of the near Earth tail and apply convection from the nightside to the dayside. We use a mesoscale part of the magnetotail extending from $x = -5 R_E$ to about $x = -50 R_E$ on the tailward side. In order to apply this convection, we need to modify the Earthward boundary to enable sunward flow in the tail region. As mentioned, such boundary conditions can be very tedious. Our system extends from $z = 0$ (equatorial plane) to $z = 20$ (lobes) where units are in Earth radii R_E , and symmetry is used such that only the northern half of the magnetotail is simulated. In the y direction, the system extends from $y = -20$ to $y = 20$ such that the boundary does not cross the magnetopause where flow becomes fast.

Note that the sunward outflow condition has to be applied not only at $z = 0$ but everywhere at our Earthward boundary. Assuming that magnetic field lines must be transported smoothly through this boundary, we actually integrate magnetic field lines from each boundary coordinate into the system and determine the location where they cross the equatorial plane. We can then assume a hypothetical smooth

convection profile in the near Earth equatorial plane and associate each field line with such a convection, i.e., map the assumed convection from the equatorial plane to the Earthward boundary. The physical motivation for this shape of convection is that magnetic flux transport should follow approximately contours of constant flux tube entropy, i.e., be mostly azimuthal. Various tests of this approach indicate that the exact details do not play an important role. Assuming that flux tubes are sort of rigidly convected, i.e., with a constant angular velocity, their azimuthal velocity is given by the mapped velocity and radial distance of the respective boundary point. Thus, it is possible to determine a unique value for the x and y velocity components of each point at the Earthward boundary. Note that the hypothetical equatorial convection is used only to compute flow at the Earthward boundary and not applied in the equatorial plane where plasma motion develops self-consistently. The assumed convection pattern and the resulting flow in the equatorial plane are illustrated in Fig. 1.2. This boundary condition for slow adiabatic convection proved quite stable and reliable to generate the desired sunward outflow. Several simulation studies have addressed this problem of a slow evolution of the near Earth tail [16–18]:

Finally, the boundary condition in the equatorial plane at $z = 0$ is an excellent example of a symmetry condition which uses the symmetry of the northern and southern hemisphere. Note that this applies only if the y component of the magnetic field in the equatorial plane is 0. The specific conditions used are symmetry $f(i_z = 1) = f(i_z = 3)$ for density ρ , pressure p magnetic field component b_z , and velocity components u_x and u_y and antisymmetry (zero derivative) $f(i_z = 1) = -f(i_z = 3)$ for the magnetic field components b_x and b_y and the velocity com-

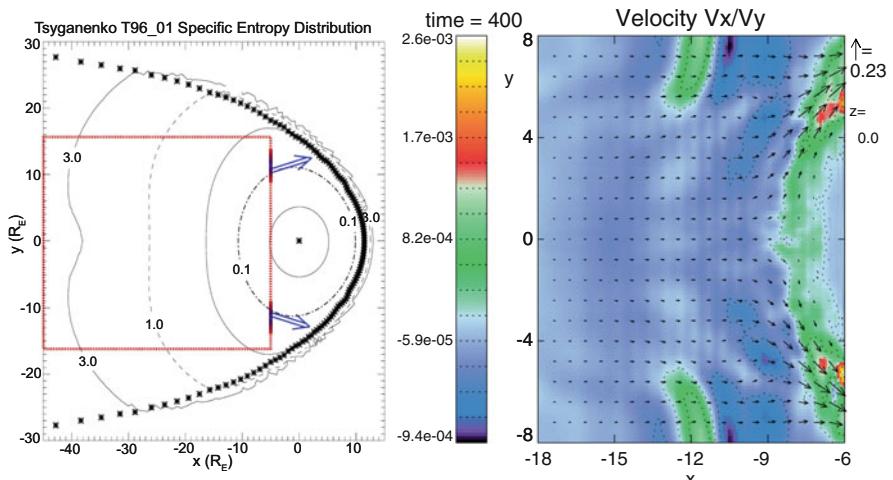


Fig. 1.2 Left, illustration of the assumed equatorial convection in order to compute flow conditions at the Earthward (right) simulation boundary indicated in red. Right, convection in the equatorial plane in response to the applied Earthward boundary condition in the simulation

ponent u_z . The remaining conditions at the lateral boundaries in y , at the lobe boundary at z_{max} , and the tailward boundary at x_{min} use symmetry $f(i = i_b + 1) = f(i = i_b - 1)$ or Dirichlet conditions and allow flow tangential to the respective boundary. The normal magnetic field component is again used to satisfy $\nabla \cdot \mathbf{b} = 0$.

1.4 Examples of MHD Simulations

The used simulation code for the following examples employs a leapfrog and DuFort-Frankel scheme using a finite difference approximation. This finite difference approximation is of second-order accuracy, and a small second-order diffusion term is applied where numerical dispersion can occur. The finite difference leapfrog approximation is also subject to a time step limitation similar to the one described in Sect. 1.2; however, it is noted that this limitation is also required for sufficient accuracy. A time step longer than allowed by the stability condition would imply a low accuracy for the propagation of the respective MHD waves.

1.4.1 Kelvin-Helmholtz Simulations

The KH instability has also the advantage that the linear evolution for sufficiently simple situations is reasonably well understood by analytic models in the limit of incompressible flow. The mode is unstable for sufficiently large flow shear compared to the stabilizing influence of the magnetic field aligned with the \mathbf{k} vector of the mode, i.e.:

$$[(\mathbf{u}_1 - \mathbf{u}_2) \cdot \mathbf{k}]^2 > \frac{\rho_1 + \rho_2}{\mu_0 \rho_1 \rho_2} [(\mathbf{u}_{A1} \cdot \mathbf{k}) + \alpha_2 (\mathbf{u}_{A2} \cdot \mathbf{k})]$$

Chandrasekhar [3] where the indices 1 and 2 indicate the two sides of the boundary and \mathbf{u}_A refers to the Alfvén speed. The setup for the initial and boundary conditions for KH simulations as described in Sect. 1.3 is quite straightforward. Therefore, the KH mode is a good test case for numerical simulations. The simulations can actually be conducted by two-dimensional and fairly low-resolution models. The results shown in Fig. 1.3 use a resolution of 203 by 203 grid points; however, a few ten grid points in each direction are sufficient to resolve the early evolution. Higher resolution allows for more details and longer accurate simulation later in the nonlinear evolution although boundaries within the nonlinear KH wave become eventually arbitrary thin within a finite time limiting also high-resolution simulations.

Figure 1.3 illustrates that the magnetic field which was initially uniform is strongly twisted within a few turnover times causing large changes of the magnetic field both in the simulation plane and perpendicular to it. Density boundaries become very thin, and the strong magnetic field deformation causes very large local current densities even for an initially uniform magnetic field. Any initial current is strongly modulated and amplified, and the nonlinear evolution can cause magnetic

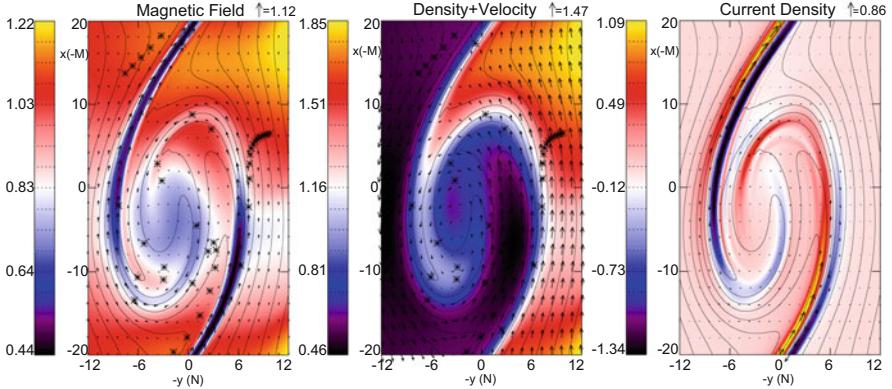


Fig. 1.3 Results from two-dimensional KH simulations. From left to right, the figures show magnetic field, density and velocity arrows, and current density

reconnection at the strong current layers within the KH vortex. More detailed results for two-dimensional KH studies and comparisons with observations can be found in the respective literature (e.g., [8, 10, 19–21]).

The nonlinear and three-dimensional evolution have provided interesting new results for the transport of plasma and magnetic flux at the magnetospheric boundaries over the past two decades. The nonlinear KH mode provides a number of different mechanisms for this transport. In a simple two-dimensional approach, the nonlinear KH wave generates a very strong current layer associated with the strong twisting of the magnetic field. Similarly, the nonlinear KH can modify and intensify a preexisting current layer. These very thin and strong current layers are easily tearing mode unstable and can undergo magnetic reconnection which causes direct plasma transport across the boundary. This very small width enables ions to easily cross these boundaries causing enhanced mixing and diffusion. Another potential mechanism is the excitation of kinetic or inertial Alfvén waves through mode conversion which again can cause significant plasma transport across such boundaries (e.g., [22]).

In three dimensions, the evolution of the KH mode depends on the variation along the direction perpendicular to the plane shown in Fig. 1.3. For instance, the magnetospheric boundary is KH unstable in the equatorial plane for strongly northward interplanetary magnetic field (IMF) as illustrated in Fig. 1.4.

However, the field geometry renders the mode stable in the northern and southern hemisphere at sufficient distance from the equatorial plane. Therefore, twisting is expected to occur only locally similar to a candy wrapper. This implies that very large currents form at the boundaries of the KH unstable regions. These currents have a predominant orientation such that the associated parallel electric field (in the presence of resistivity) tends to decouple the magnetic field as illustrated in Fig. 1.4. It has been demonstrated that this provides a powerful transport mechanism that can quantitatively explain the plasma transport during extended periods of northward [23] and southward IMF [12, 24].

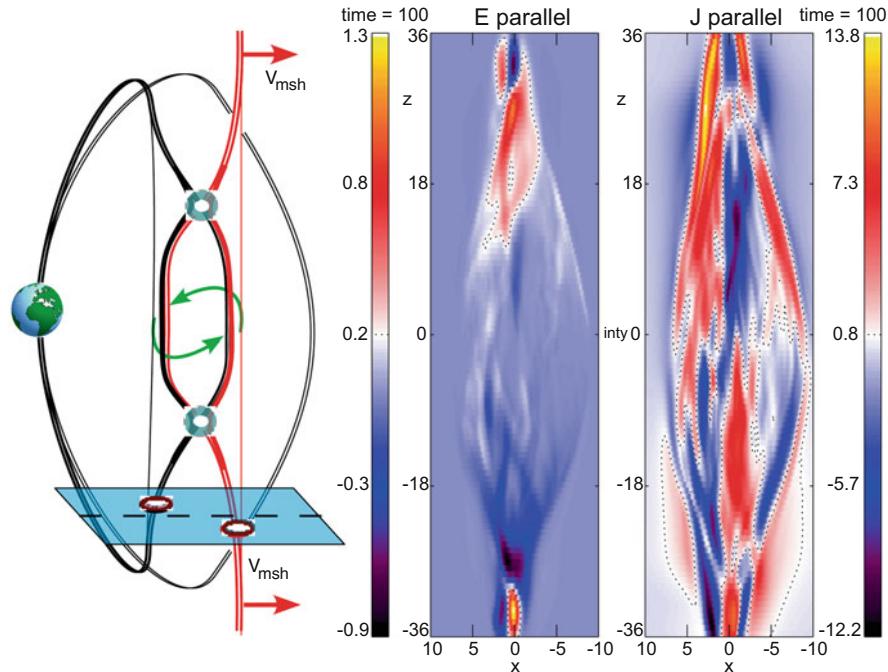


Fig. 1.4 Sketch of the KH-associated magnetic field twisting in the equatorial (left) and parallel current and electric field for a locally unstable KH wave (integrated along the direction of the velocity shear)

Recent new satellite observations from high temporal resolution particle measurements by the magnetospheric multi-scale mission (MMS) indeed show evidence of multiple reconnection processes at locations consistent with the reconnection sites from simulations, i.e., close to the equatorial plane, and in the northern and southern hemispheres [25, 26]. Nonlinear KH modes may also be the dominant instability mechanisms at the boundaries of the fast rotating giant magnetospheres [27, 28].

1.4.2 Simulations of Sunward Moving Plasma at the Earth's Bow Shock

Although KH simulations can include some shock physics for sufficiently fast jetting plasma, typically shocks associated with KH modes are weak. However, shock physics is important for bow shock transients such as HFAs. For many applications such as KH simulations or slow magnetotail evolution, the MHD equations can be closed using an energy equation like the form of Eq.(1.13). This has the additional advantage that entropy is explicitly conserved by this formulation in ideal MHD. However, for the fast shock simulations illustrated here,

the MHD equations must use total energy conservation presented in Eq.(1.12) because otherwise properties of strong shocks will not be correctly addressed in the simulation. This also applies to slow shocks in cases of very low plasma β .

HFA events upstream of the Earth's bow shock are characterized by a very hot plasma that is streaming much slower than the solar wind or sometimes has a velocity component toward the sun. Originally, HFAs were addressed as diamagnetic cavities [29–32] and believed to be associated with a tangential discontinuity (TD; TDs are transition layers in the plasma across with no particle transport across them) in the interplanetary magnetic field (e.g., [33, 34]). This was expressed by the original terminology “active current sheet” but found support also through early hybrid simulation studies [13, 35]. However, frequently HFAs can occur without the presence of a TD, and often HFAs do not show a motional electric field toward the TD as predicted by hybrid simulations [36–38].

One of the most surprising properties of HFAs is their strong deflection of the solar wind momentum. In shock physics, such a deflection can easily occur if upstream conditions change this momentum, for instance, through a significantly lower density. Lower-density regions are actually observed and associated with the foreshock regions. These density depletions are termed foreshock cavities. The example in Fig. 1.5 illustrates the evolution if a low-density flux tube interacts with a fast shock. Since the fast-mode speed is much higher in the low-density medium, the effective Mach number based on the original shock conditions is much lower for the low-density region. This implies that there must a new shock forming, which moves rapidly into the upstream direction as illustrated by the results shown in Fig. 1.5.

The resulting structure is an excellent example of shock physics. At the leading edge, the new fast shock is heating the material to temperatures higher than the magnetosheath. In the new structure, the pressure is higher than ion the solar wind but lower than in the original magnetosheath. Therefore, the structure expands into the solar wind yielding a wedge-like shape. Material from the magnetosheath is streaming into the structure from behind as illustrated in the sketch in Fig. 1.6. The figure also illustrates the evolution of strong current sheets in the newly formed structure and regions of different entropy. Note that the different colored entropy regions are all resulting from fast shocks. The upstream region just represents the original solar wind and the density-depleted flux tube. The downstream regions have the original magnetosheath (colored in cyan), the shocked plasma at the forward edge of the wedge (yellow), and two shoulders with a moderately increased entropy (red), because the expansion into the solar wind generates also fast shocks.

The structure in Fig. 1.5 appears stationary because the simulation is in the de Hoffmann-Teller frame where the electric field is 0, i.e., convection is magnetic field-aligned. In the normal incident frame, the structure is moving into the negative z direction with a velocity that depends on θ_{bn} , i.e., the angle of the magnetic field with the shock normal and on the magnitude of the solar wind (upstream velocity). It is worth noting that these simulation results suggest larger and slower moving structures for small θ_{bn} , i.e., quasiparallel shock conditions.

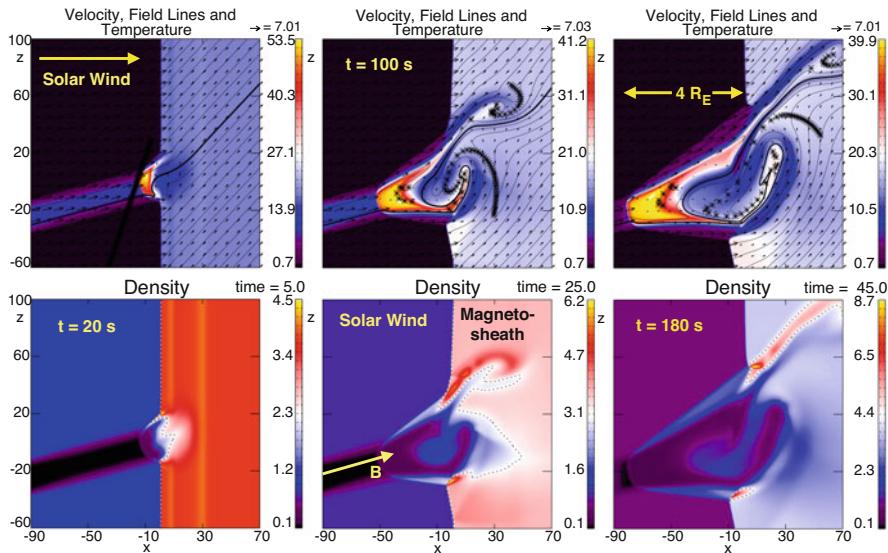


Fig. 1.5 Time sequence of the interaction of a density-depleted magnetic flux tube with a fast shock. The top plots show velocity (arrows), temperature (color), and magnetic field (lines) and the bottom plots show density

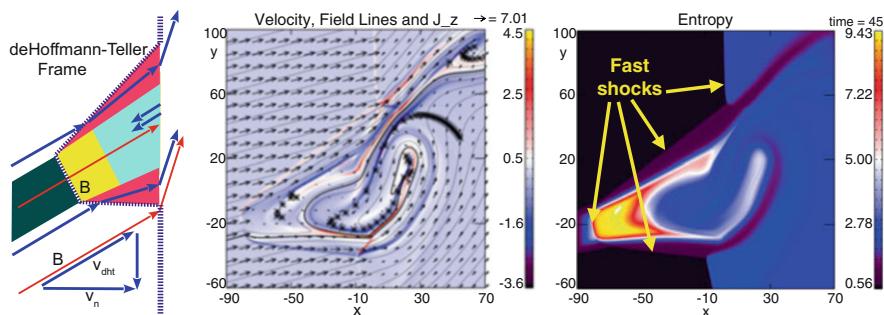


Fig. 1.6 From left to right, the plots show a sketch of the newly formed shock structure caused by the low-density flux tube, plot of the velocity (arrows) current density (color) and field lines of the newly formed structure, and plot of the entropy as defined as $p^{1/\gamma} / \rho$

Note, however, that particularly for parallel conditions, MHD is not a good approximation for fast shocks. Nevertheless, one would expect that some basic properties from the MHD simulations should be observable. In any case, it is remarkable that the MHD simulations seem to capture a number of observed HFA properties quite well.

1.4.3 Earth's Magnetotail Simulations

The final example addresses the slow evolution of a three-dimensional system that is close to equilibrium sometime also addressed as quasi-static evolution indicating that the system satisfies force balance at all times. This type of evolution is important to understand the transition from a stable configuration to an unstable state. Examples for such an evolution include solar prominences and, in the present case, the magnetotail during the growth phase of a substorm [39–41]. It is well established that the onset of a substorm involves very thin current layers in the near Earth tail [42–44]. These current sheets can be generated by various mechanisms, for instance, instabilities and other dynamic processes as illustrated by the prior examples. In terms of a quasi-static evolution, it has been suggested that a slow compression of the lobes of the magnetotail can cause thin current sheet formation. Boundary conditions for this type of evolution play a somewhat different role than in other examples. While in other examples boundary conditions just need to be consistent with the initial conditions and the evolutions, here, boundary conditions provide a major constraint and a slow driver for the change of the equilibrium. So, in this example, boundary conditions cause the evolution of the system which is slow up to the point where some critical state is reached.

The simulation results in Fig. 1.7 use a magnetic flux transport rate of about 50 kV well comparable to a typical cross-polar cap potential during the growth phase of a substorm. The initial magnetic field B_z component in the equatorial plane represents a typical magnetotail field because it is based on the Tsyganenko model magnetic field [45]. During 50 min of sunward convection, out of the near Earth magnetotail, the closed magnetic field in the near Earth region is strongly reduced in some regions to values below 1 nT. The larger values close to the Earthward boundary at $z = -6R_E$ are close to 100 nT initially and are reduced to about 60 nT

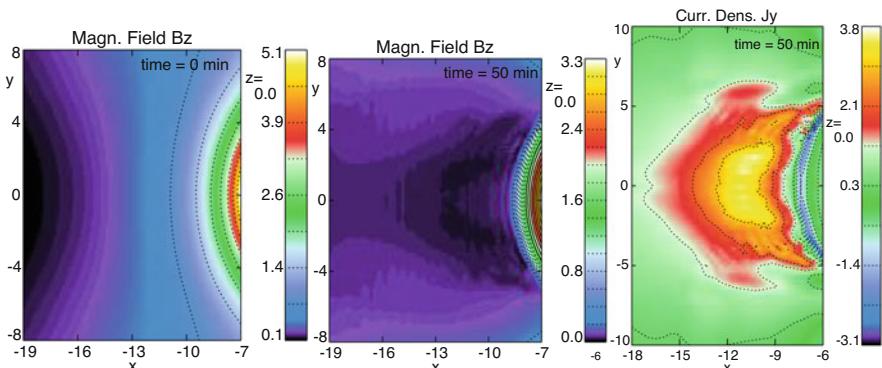


Fig. 1.7 Evolution of the magnetic field component B_z in the near Earth equatorial plane shown at times 0 and 50 min into the evolution and cross-tail current density component j_y at time = 50 min. In the chosen normalization, length units correspond to 1 Earth radius, and magnetic field is normalized to 20 nT

after 50 min of slow convection. The last plot in Fig. 1.7 presents the cross-tail current density component in the equatorial plane at time = 50 min. It illustrates a concentrated and amplified current density in the region where the magnetic field B_z is reduced.

In order to demonstrate that the evolution is quasi-static, i.e., a sequence of equilibrium configurations, we have run tests with higher and lower values of the magnetic flux transport rate. The evolution was almost identical but faster for higher potential and slower for lower potential. The state of the near Earth magnetosphere was determined by the amount of magnetic flux that was removed from the region and not by the flux removal rate or by the specific time of the evolution. More details and other results on this type of simulation are provided by Hsieh and Otto [16, 17] and Otto et al. [18]. Note that very similar results have been obtained using the rice convection model [46] and global simulations [47].

Figure 1.8 presents the cross-tail current density in the midnight meridian in order to demonstrate that the quasi-static evolution in this example indeed leads to very thin and strong current layers. The current density in this region is as high as 10 nA/m^2 and about an order of magnitude larger than typical equilibrium current densities. The process that is illustrated here has been termed current

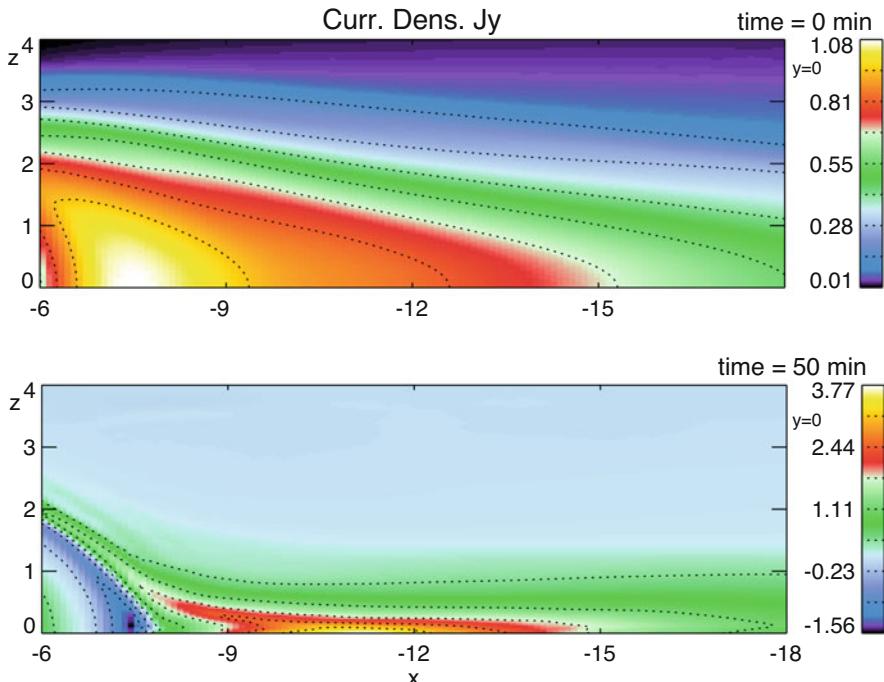


Fig. 1.8 Illustration of the cross-tail current density at times 0 and 50 min corresponding to the times shown in Fig. 1.7 for the equatorial B_z

sheet formation through magnetic flux depletion in order to distinguish it from the typical assumption that current sheet formation is caused by lobe compression due to additional magnetic flux in the lobe region of the magnetotail. Note that during a typical growth phase, both processes should operate. Dayside reconnection generates open magnetic flux that is convected to the nightside and added to the lobes. Dayside reconnection also causes a reduction of flux on the dayside that is compensated by convection from the night side as can be seen from typical polar cap convection patterns. We also note that observations of current sheets agree well with the presented model results [48–50].

1.5 Summary and Conclusions

The examples illustrated here demonstrate various aspects of numeric simulations of the MHD equations. It is emphasized that all numerical solutions to the MHD equations have errors that in many cases can be addressed as numerical diffusion or dispersion. These errors limit the validity of numerical simulations, and in general any application has to insure that numerical errors are sufficiently small in order to identify the important physics in a numerical simulation. We have illustrated a number of tests such as magnetic line conservation as well as mass, momentum, energy, and entropy conservation. The conservation laws can be employed in different ways to test numerical MHD simulations. Further tests can be conducted using known analytic properties, such as the growth rate of instabilities, e.g., the KH wave, or the jump conditions associated with fast or slow shocks. Note that this approach is somewhat different from tests of global simulations which frequently use observed magnetospheric states and ask the question of how well a given simulation can reproduce the observed properties.

Local or mesoscale MHD simulations are frequently used to examine clearly defined local processes and instabilities. They have the advantage of providing much higher resolution and are ideally suited to examine a sensitive dependence on plasma or boundary conditions for a physical process. Because of their much higher resolution, local models often provide better conservation laws than global processes where by necessity a KH mode or magnetic reconnection is only resolved by a very limited number of grid points or nodes. Typical applications of local MHD but also kinetic models frequently address processes at the magnetospheric boundaries, at the bow shock, and in the magnetotail of planetary magnetospheres similar to the examples and references provided in this chapter. Similar local or mesoscale simulations are used in solar and astrophysical simulations (e.g., [51–53]).

Usually, it is assumed that higher-order approximations are better than low-order approximations. This is in general true, particularly, if higher-order derivatives are important in the basic equations. However, this advantage truly applies only if the grid resolution is similar for higher- and lower-order approximations. We note that first-order approximations are generally not acceptable because of the large numerical diffusion. Given that it is more complicated to formulate a higher-order scheme

and that it is much more involved to formulate consistent boundary conditions, it really depends on the application which type of scheme is advantageous.

A similar remark applies to implicit schemes. Implicit schemes are great because they may remove or reduce time step limitations and can improve stability. However, it should also be kept in mind that the choice of a time step must also consider accuracy, and for many applications, accuracy and stability imply a very similar choice of the time step for explicit codes. Numerical techniques in space physics include a large number of finite difference (e.g., the examples and references provided here) and finite volume methods (e.g., [47, 54, 55]). In contrast, finite element seems to play no role in space plasma MHD simulation, and spectral or pseudo-spectral methods are relatively rare (e.g., [56]). Finite element methods have the advantage that they are particularly useful to model more complicated shapes of obstacles or the system domain. Spectral methods suffer from the difficulty of the formulation of boundary conditions which need to be consistent with the chosen base functions. The difficulty of formulating boundary conditions has been illustrated particularly by our example 3.

Finally, numerical simulations are sometimes used like a black box, i.e., without consideration of the numerical limitations for a code in a given application. Clearly, when this happens, there can be little confidence in the results. Any applications of numerical simulations should keep in mind the capabilities of the simulation method and use consistency checks to generate confidence in the results.

Further reading and additional information on MHD simulations and methods are provided in Chap. 7 titled “Adaptive Global Magnetohydrodynamics Simulations” and Chap. 12 “Higher-Order Magnetohydrodynamic Numerics.”

References

1. N.A. Krall, A.W. Trivelpiece, *Principles of Plasma Physics* (San Francisco Press, San Francisco, 1986)
2. H. Alfvén, C.G. Fälthammar, *Cosmical Electrodynamics* (Oxford University Press, Oxford, 1963)
3. S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability* (Oxford University Press, New York, 1961)
4. E.R. Priest, *Solar Magnetohydrodynamics* (D. Reidel Publishing, Dordrecht, 1987)
5. D.E. Potter, *Computational Physics* (Wiley, New York, 1973)
6. C.A.J. Fletcher, *Computational Techniques for Fluid Dynamics*, vol. 1 (Springer, Berlin, 1991)
7. A. Otto, Phys. Fluids B **3**, 1739 (1991)
8. A. Otto, D.H. Fairfield, J. Geophys. Res. **105**, 21 (2000). <https://doi.org/10.1029/1999JA000312>
9. A. Miura, Phys. Rev. Lett. **49**, 779 (1982)
10. K. Nykyri, A. Otto, Geophys. Res. Lett. **28**, 3565 (2001). <https://doi.org/10.1029/2001GL013239>
11. H. Hasegawa, M. Fujimoto, T.D. Phan, H. Rème, A. Balogh, M.W. Dunlop, C. Hashimoto, R. TanDokoro, Nature **430**, 755 (2004). <https://doi.org/10.1038/nature02799>
12. X. Ma, A. Otto, P.A. Delamere, J. Geophys. Res. **119**, 781 (2014). <https://doi.org/10.1002/2013JA019224>

13. V.A. Thomas, D. Winske, M.F. Thomsen, T.G. Onsager, *J. Geophys. Res.* **96**, 11 (1991). <https://doi.org/10.1029/91JA01092>
14. Y. Lin, *J. Geophys. Res.* **102**, 24265 (1997). <https://doi.org/10.1029/97JA01989>
15. N. Omidi, D.G. Sibeck, *J. Geophys. Res.* **112**, A01203 (2007). <https://doi.org/10.1029/2006JA011663>
16. M.S. Hsieh, A. Otto, *J. Geophys. Res.* **119**, 3430 (2014). <https://doi.org/10.1002/2013JA019459>
17. M.S. Hsieh, A. Otto, *J. Geophys. Res.* **120**, 4264 (2015). <https://doi.org/10.1002/2014JA020925>
18. A. Otto, M.S. Hsieh, F. Hall, *Current Sheets Formation in Planetary Magnetotail* (Wiley, New York, 2015), pp. 289–305. *Geophys. Monogr., Magnetotails in the Solar System*. <https://doi.org/10.1002/9781118842324.ch17>
19. A. Miura, *J. Geophys. Res.* **89**, 801 (1984)
20. K. Nykyri, A. Otto, B. Lavraud, C. Mouikis, L.M. Kistler, A. Balogh, H. Rème, *Ann. Geophys.* **24**, 2619 (2006). <https://doi.org/10.5194/angeo-24-2619-2006>
21. T.K.M. Nakamura, M. Fujimoto, A. Otto, *Geophys. Res. Lett.* **33**, 14106 (2006). <https://doi.org/10.1029/2006GL026318>
22. S. Wing, J.R. Johnson, C.C. Chaston, M. Echim, C.P. Escoubet, B. Lavraud, C. Lemon, K. Nykyri, A. Otto, J. Raeder, C.P. Wang, *Space Sci. Rev.* **184**, 33 (2014). <https://doi.org/10.1007/s11214-014-0108-9>
23. X. Ma, P. Delamere, A. Otto, B. Burkholder, *J. Geophys. Res. (Space Physics)* **122**(A11), 10 (2017). <https://doi.org/10.1002/2017JA024394>
24. X. Ma, A. Otto, P.A. Delamere, *J. Geophys. Res.* **119**, 808 (2014). <https://doi.org/10.1002/2013JA019225>
25. S. Eriksson, B. Lavraud, F.D. Wilder, J.E. Stawarz, B.L. Giles, J.L. Burch, W. Baumjohann, R.E. Ergun, P.A. Lindqvist, W. Magnes, C.J. Pollock, C.T. Russell, Y. Saito, R.J. Strangeway, R.B. Torbert, D.J. Gershman, Y.V. Khotyaintsev, J.C. Dorelli, S.J. Schwartz, L. Avanov, E. Grimes, Y. Vernisse, A.P. Sturner, T.D. Phan, G.T. Marklund, T.E. Moore, W.R. Paterson, K.A. Goodrich, *Geophys. Res. Lett.* **43**, 5606 (2016). <https://doi.org/10.1002/2016GL068783>
26. W. Li, M. André, Y.V. Khotyaintsev, A. Vaivads, D.B. Graham, S. Toledo-Redondo, C. Norrgren, P. Henri, C. Wang, B.B. Tang, B. Lavraud, Y. Vernisse, D.L. Turner, J. Burch, R. Torbert, W. Magnes, C.T. Russell, J.B. Blake, B. Mauk, B. Giles, C. Pollock, J. Fennell, A. Jaynes, L.A. Avanov, J.C. Dorelli, D.J. Gershman, W.R. Paterson, Y. Saito, R.J. Strangeway, *Geophys. Res. Lett.* **43**, 5635 (2016). <https://doi.org/10.1002/2016GL069192>
27. X. Ma, B. Stauffer, P.A. Delamere, A. Otto, *J. Geophys. Res. (Space Physics)* **120**, 1867 (2015). <https://doi.org/10.1002/2014JA020746>
28. B. Zhang, P.A. Delamere, X. Ma, B. Burkholder, M. Wilberger, J.G. Lyon, V.G. Merkin, K.A. Sorathia, *Geophys. Res. Lett.* **45**, 56 (2018). <https://doi.org/10.1002/2017GL076315>
29. S.J. Schwartz, C.P. Chaloner, P.J. Christiansen, A.J. Coates, D.S. Hall, A.D. Johnstone, M.P. Gough, A.J. Norris, R.P. Rijnbeek, D.J. Southwood, L.J.C. Woolliscroft, *Nature* **318**, 269 (1985). <https://doi.org/10.1038/318269a0>
30. M.F. Thomsen, J.T. Gosling, S.A. Fuselier, S.J. Bame, C.T. Russell, *J. Geophys. Res.* **91**, 2961 (1986). <https://doi.org/10.1029/JA091iA03p02961>
31. M.F. Thomsen, J.T. Gosling, S.J. Bame, K.B. Quest, C.T. Russell, S.A. Fuselier, *J. Geophys. Res.* **93**, 11311 (1988). <https://doi.org/10.1029/JA093iA10p11311>
32. G. Paschmann, G. Haerendel, N. Sckopke, E. Möbius, H. Luehr, C.W. Carlson, *J. Geophys. Res.* **93**, 11279 (1988). <https://doi.org/10.1029/JA093iA10p11279>
33. S.J. Schwartz, R.L. Kessel, C.C. Brown, L.J.C. Woolliscroft, M.W. Dunlop, C.J. Farrugia, D.S. Hall, *J. Geophys. Res.* **93**, 11295 (1988). <https://doi.org/10.1029/JA093iA10p11295>
34. S.J. Schwartz, *Adv. Space Res.* **15**, 107 (1995). [https://doi.org/10.1016/0273-1177\(95\)00025-A](https://doi.org/10.1016/0273-1177(95)00025-A)
35. V.A. Thomas, S.H. Brecht, *J. Geophys. Res.* **93**, 11341 (1988). <https://doi.org/10.1029/JA093iA10p11341>

-
36. H. Zhang, D.G. Sibeck, Q.G. Zong, N. Omidi, D. Turner, L.B.N. Clausen, J. Geophys. Res. **118**, 3357 (2013). <https://doi.org/10.1002/jgra.50376>
 37. L.L. Zhao, Q.G. Zong, H. Zhang, S. Wang, J. Geophys. Res. **120**, 6332 (2015). <https://doi.org/10.1002/2014JA020862>
 38. L.L. Zhao, H. Zhang, Q.G. Zong, J. Geophys. Res. **122**, 235 (2017). <https://doi.org/10.1002/2016JA023319>
 39. K. Schindler, J. Birn, J. Geophys. Res. **87**, 2263 (1982)
 40. K. Schindler, J. Birn, J. Geophys. Res. **98**, 15477 (1993). <https://doi.org/10.1029/93JA01047>
 41. J. Birn, M. Hesse, K. Schindler, J. Geophys. Res. **103**, 6843 (1998). <https://doi.org/10.1029/97JA03602>
 42. V.A. Sergeev, P. Tanskanen, K. Mursula, A. Korth, R.C. Elphic, J. Geophys. Res. **95**, 3819 (1990). <https://doi.org/10.1029/JA095iA04p03819>
 43. S. Ohtani, K. Takahashi, L.J. Zanetti, T.A. Potemra, R.W. McEntire, T. Iijima, J. Geophys. Res. **97**, 19311 (1992)
 44. V.A. Sergeev, A.V. Nikolaev, M.V. Kubyshkina, N.A. Tsyganenko, H.J. Singer, J.V. Rodriguez, V. Angelopoulos, R. Nakamura, S.E. Milan, J.C. Coxon, B.J. Anderson, H. Korth, J. Geophys. Res. **119**, 9714 (2014). <https://doi.org/10.1002/2014JA020522>
 45. N.A. Tsyganenko, D.P. Stern, J. Geophys. Res. **101**, 27187 (1996)
 46. X.Z. Zhou, V. Angelopoulos, A. Runov, M.I. Sitnov, F. Coroniti, P. Pritchett, Z.Y. Pu, Q.G. Zong, J.P. McFadden, D. Larson, K.H. Glassmeier, J. Geophys. Res. **114**(A13), A03223 (2009). <https://doi.org/10.1029/2008JA013777>
 47. E. Gordeev, V. Sergeev, V. Merkin, M. Kuznetsova, Geophys. Res. Lett. **44**, 8696 (2017). <https://doi.org/10.1002/2017GL074539>
 48. A.A. Petrukovich, W. Baumjohann, R. Nakamura, H. Rème, J. Geophys. Res. **114**(A13), A09203 (2009). <https://doi.org/10.1029/2009JA014064>
 49. A. Petrukovich, A. Artemyev, I. Vasko, R. Nakamura, L. Zelenyi, Space Sci. Rev. (2015). <https://doi.org/10.1007/s11214-014-0126-7>
 50. A.V. Artemyev, V. Angelopoulos, A. Runov, A.A. Petrukovich, J. Geophys. Res. **121**, 6718 (2016). <https://doi.org/10.1002/2016JA022779>
 51. A. Otto, J. Büchner, B. Nikutowski, Astron. Astrophys. **468**, 313 (2007). <https://doi.org/10.1051/0004-6361:20054495>
 52. J.C. Santos, J. Büchner, A. Otto, Astron. Astrophys. **525**, A3+ (2011). <https://doi.org/10.1051/0004-6361/201014758>
 53. E. Adamson, J. Büchner, A. Otto, Astron. Astrophys. **557**, A118 (2013). <https://doi.org/10.1051/0004-6361/201321281>
 54. V.G. Merkin, M.I. Sitnov, J.G. Lyon, J. Geophys. Res. **120**, 1993 (2015). <https://doi.org/10.1002/2014JA020651>
 55. B. Zhang, K.A. Sorathia, J.G. Lyon, V.G. Merkin, J.S. Garretson, M. Wiltberger, Astrophys. J. Suppl. Ser. **244**(1), 20 (2019). <https://doi.org/10.3847/1538-4365/ab3a4c>
 56. H.C. Ku, D.G. Sibeck, J. Geophys. Res. **102**(A2), 2243 (1997). <https://doi.org/10.1029/96JA03162>



Hall Magnetohydrodynamics

2

J. D. Huba

Abstract

An overview of Hall magnetohydrodynamic (MHD) physics is presented. The theoretical underpinnings of Hall MHD are described followed by a discussion of numerical methods used to model Hall MHD processes. These methods are highlighted by applying them to physical systems that are affected by Hall physics such as sub-Alfvénic plasma expansions, plasma opening switches, and magnetic reconnection.

2.1 Introduction

Hall magnetohydrodynamic (MHD) theory has been used to describe and understand a variety of interesting space and laboratory plasma phenomena over the last three decades. Hall MHD introduces additional physical processes on scale sizes smaller than the ion inertial length (but larger than electron scales) such as new wave modes and instabilities, enhanced magnetic field transport, and fast magnetic reconnection rates. The theory has been successfully applied to the structuring of sub-Alfvénic plasma expansions [1–8] and to rapid magnetic field transport in plasma opening switches [9–18]. It was suggested that Hall physics plays a critical role in magnetic reconnection physics [19–40] and this has spurred renewed interest in this subject [41–44]. 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 update an earlier tutorial on the subject [45]. The overview of the underlying physics associated with

J. D. Huba (✉)

Syntek Technologies, Cheverly, MD, USA

e-mail: jdhuba@syntek.org

Hall physics remains the same but has added a short section on implicit numerical techniques. Additionally, we update the application to magnetic reconnection, discuss the structuring associated with sub-Alfvénic plasma expansions, and discuss the application of Hall MHD to global modeling of magnetospheres.

2.2 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 [46]:

$$\frac{m_e}{ne^2} \frac{\partial \mathbf{J}}{\partial t} - \frac{1}{ne} \nabla \cdot \overline{\overline{P}_e} = \mathbf{E} + \frac{1}{c} \mathbf{V}_i \times \mathbf{B} - \frac{1}{nec} \mathbf{J} \times \mathbf{B} - \mathbf{J}/\sigma \quad (2.1)$$

where σ is the plasma conductivity. We simplify (2.1) by assuming $L \gg c/\omega_{pe}$, $L \gg \rho_e$, and $\sigma \rightarrow \infty$ where ω_{pe} is the electron plasma frequency, ρ_e is the 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:

$$\mathbf{E} = -\frac{1}{c} \mathbf{V}_i \times \mathbf{B} + \underbrace{\frac{1}{nec} \mathbf{J} \times \mathbf{B}}_{\text{Hall term}}. \quad (2.2)$$

The final term in (2.2) is identified as the Hall term. Physically, the Hall term decouples ion and electron motion on ion inertial length scales: $L \lesssim c/\omega_{pi}$ where $\omega_{pi} = 4\pi n_i e^2 / m_i$.

The full set of Hall MHD equations is as follows:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \mathbf{V} = 0 \quad (2.3)$$

$$\frac{\partial \rho \mathbf{V}}{\partial t} + \nabla \cdot \left[\rho \mathbf{V} \mathbf{V} + (P + B^2/8\pi) \underline{\underline{\mathbf{I}}} - \mathbf{B} \mathbf{B} / 4\pi \right] = 0 \quad (2.4)$$

$$\frac{\partial \epsilon}{\partial t} + \nabla \cdot \left[\mathbf{V} (\epsilon + P + B^2/8\pi) - \mathbf{B} / 4\pi (\mathbf{V} \cdot \mathbf{B}) \right] +$$

$$\nabla \cdot \left[\mathbf{V}_H (B^2/8\pi) - \mathbf{B} / 4\pi (\mathbf{V}_H \cdot \mathbf{B}) \right] = 0 \quad (2.5)$$

$$\frac{\partial \mathbf{B}}{\partial t} = -c \nabla \times \mathbf{E} = \nabla \times [(\mathbf{V} + \mathbf{V}_H) \times \mathbf{B}] \quad (2.6)$$

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

$$P = (\gamma - 1)(\epsilon - \rho V^2/2 - B^2/8\pi). \quad (2.7)$$

We take $\gamma = 5/3$. In writing (2.3)–(2.6), 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 (2.6), the Hall term only enters the energy equation (2.5). 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 (2.5) is to use the adiabatic pressure equation:

$$\frac{\partial P}{\partial t} + \nabla \cdot P \mathbf{V} = -(\gamma - 1)P \nabla \cdot \mathbf{V}. \quad (2.8)$$

However, a computational problem arises using (2.8): 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:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{V}_H \times \mathbf{B}) = -\nabla \times (\mathbf{J}/ne \times \mathbf{B}). \quad (2.9)$$

We rewrite (2.9) as follows to identify the origin of each wave mode:

$$\frac{\partial \mathbf{B}}{\partial t} = \underbrace{-\frac{1}{ne} \nabla \times (\mathbf{J} \times \mathbf{B})}_{\text{Whistler waves}} + \underbrace{\frac{1}{n^2 e} \nabla n \times (\mathbf{J} \times \mathbf{B})}_{\text{Hall drift waves}}. \quad (2.10)$$

The first term on the RHS of (2.10) 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 details.

2.2.1 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:

$$\frac{\partial \delta \mathbf{B}}{\partial t} = -\frac{1}{ne} \nabla \times (\delta \mathbf{J} \times \mathbf{B}) . \quad (2.11)$$

We linearize (2.11) to obtain:

$$\omega \delta \mathbf{B} = \frac{1}{ne} (k_z B \delta J_x \hat{\mathbf{e}}_x + k_z B \delta J_y \hat{\mathbf{e}}_y) . \quad (2.12)$$

Making use of Ampere's law (i.e., $\delta \mathbf{J} = (c/4\pi) \nabla \times \delta \mathbf{B}$), we obtain the coupled equations:

$$\omega \delta B_x = -i \frac{ck_z^2 B}{4\pi ne} \delta B_y \quad (2.13)$$

$$\omega \delta B_y = i \frac{ck_z^2 B}{4\pi ne} \delta B_x . \quad (2.14)$$

Finally, solving (2.13) and (2.14), the whistler wave dispersion relation is obtained:

$$\omega = \frac{k_z^2 B}{4\pi ne} = k_z V_A \left(\frac{k_z c}{\omega_{pi}} \right) . \quad (2.15)$$

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.

2.2.2 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:

$$\frac{\partial \delta \mathbf{B}}{\partial t} = \frac{1}{n^2 e} \nabla n \times (\delta \mathbf{J} \times \mathbf{B}) . \quad (2.16)$$

We linearize (2.16) and find that

$$i\omega\delta B_z = \frac{1}{n^2 e} \frac{\partial n}{\partial x} \delta J_x B . \quad (2.17)$$

Noting that $\delta J_x = ik_y(c/4\pi)\delta B_z$, the following dispersion relation is obtained from (2.17):

$$\omega = \frac{k_y B}{4\pi n e} \frac{1}{n} \frac{\partial n}{\partial x} = k_y V_A \left(\frac{c}{L_n \omega_{pi}} \right) \quad (2.18)$$

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 (2.18) 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 [12, 47].

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 (2.9) [48, 49]. Rudakov and Huba [50] found that a rapid, localized thinning of a current layer supported by a unidirectional 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.

2.3 Numerical Methods

We describe in some detail the numerical techniques used in the NRL 3D Hall MHD code VooDoo to solve (2.3)–(2.6). 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 sub-cycling the Hall term for computational efficiency.

2.3.1 Cell Definition

The code uses a Cartesian, staggered mesh, known as the Yee grid, and is shown in Fig. 2.1. 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.

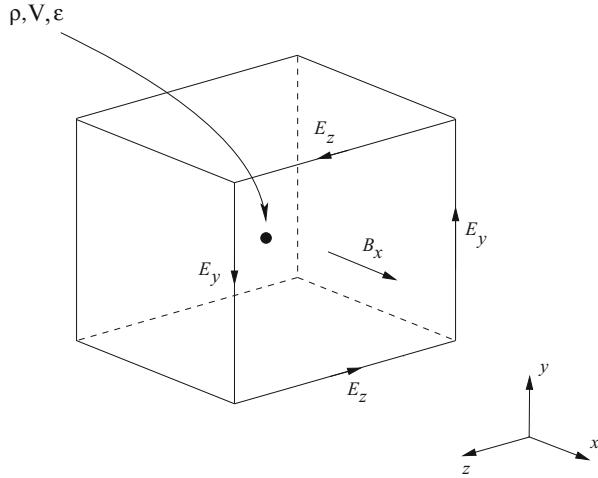


Fig. 2.1 Mesh used in the NRL 3D Hall code VooDoo

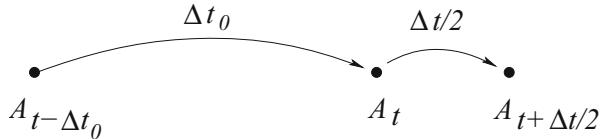


Fig. 2.2 Schematic of the time step scheme to obtain variables at the half time step

2.3.2 Time Step Scheme

The code uses a second-order Adams-Bashforth time-stepping scheme. The time advancement scheme to update the variable A from time t to time $t + \Delta t$ is:

$$A^{t+\Delta t} = A^t + \Delta t \mathcal{F}_A^{t+\Delta t/2} \quad (2.19)$$

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 \mathcal{F} are obtained by a linear extrapolation from the two previous time steps. Specifically, the half time step values are given by:

$$A^{t+\Delta t/2} = A^t + \frac{\Delta t}{2\Delta t_0} (A^t - A^{t-\Delta t_0}) \quad (2.20)$$

where the time steps are shown schematically in Fig. 2.2. The code uses a variable time step so the full time step increments (Δt_0 and Δt) can be different.

2.3.3 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:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \rho \mathbf{V}. \quad (2.21)$$

This equation is integrated over a cell volume:

$$\int \frac{\partial \rho}{\partial t} d^3x = - \int \nabla \cdot \rho \mathbf{V} d^3x \quad (2.22)$$

which can be written as:

$$\frac{\partial \rho_T}{\partial t} = - \oint (\rho \mathbf{V}) \cdot \hat{\mathbf{n}} d^2x \quad (2.23)$$

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:

$$\rho_T^{t+\Delta t} = \rho_T^t - \Delta t \oint_{\text{flux } \mathbf{F}} \underbrace{(\rho \mathbf{V})^{t+\Delta t/2}}_{\text{flux } \mathbf{F}} \cdot \hat{\mathbf{n}} d^2x \quad (2.24)$$

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:

$$A^{t+\Delta t} = A^t - \Delta t \int \mathbf{F}_A^{t+\Delta t/2} \cdot \mathbf{n} d^2x \quad (2.25)$$

where \mathbf{F}_A is the flux of A .

2.3.4 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 Ref. [51] 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.

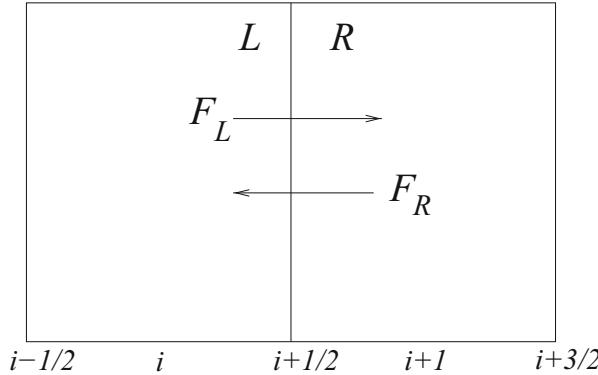


Fig. 2.3 Flux components F_L and F_R at a cell face $i + 1/2$

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:

$$F = \rho V = \rho \int dv v f(v) \quad (2.26)$$

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:

$$F_{i+1/2} = (\rho V)_{i+1/2} = (\rho V)_{i+1/2}^L + (\rho V)_{i+1/2}^R \quad (2.27)$$

where $F_L = (\rho V)_{i+1/2}^L$ and $F_R = (\rho V)_{i+1/2}^R$ in Fig. 2.3. 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:

$$F_L = \rho^L \int_0^\infty dv v f^L(v) \quad (2.28)$$

and

$$F_R = \rho^R \int_{-\infty}^0 dv v f^R(v) \quad (2.29)$$

so that the total flux at a cell interface is:

$$F = F_R + F_L . \quad (2.30)$$

In (2.28), only particles moving in the positive direction contribute to the flux F_L ; hence, the integration limits are from 0 to ∞ . Conversely, in (2.29), 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.

2.3.4.1 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 [52]: 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., 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. 2.4. We now define the function $\mathcal{G}(x)$ as:

$$\mathcal{G}_c(x) = \int_{-\infty}^x G_c(s) ds . \quad (2.31)$$

From (2.31), it follows that the high-order value is:

$$G_c(x) = \frac{\partial \mathcal{G}_c}{\partial x} . \quad (2.32)$$

We apply (2.32) to find G_c to second-order at the interface $i + 1/2$ and find that:

$$G_{c(2)}^{i+1/2} = \left(\frac{\partial \mathcal{G}_c}{\partial x} \right)_{i+1/2} = \frac{\mathcal{G}_c^{i+3/2} - \mathcal{G}_c^{i-1/2}}{2\Delta x} . \quad (2.33)$$

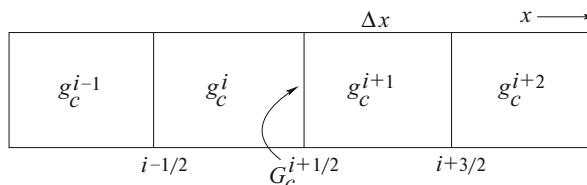


Fig. 2.4 One-dimensional grid for the high-order interpolation scheme

Now, we note that the meaning of the last term in (2.33) is:

$$\begin{aligned} \frac{\mathcal{G}_c^{i+3/2} - \mathcal{G}_c^{i-1/2}}{2\Delta x} &= \frac{1}{2\Delta x} [\text{integral of cells } i \text{ and } i + 1] \\ &= \frac{1}{2\Delta x} (g_c^{i+1} + g_c^i) \Delta x \end{aligned} \quad (2.34)$$

so that:

$$G_{c(2)}^{i+1/2} = \frac{1}{2} (g_c^{i+1} + g_c^i). \quad (2.35)$$

This technique can be applied to obtain higher-order interpolated values of G_c . For example, to fourth order:

$$G_{c(4)}^{i+1/2} = \frac{1}{12\Delta x} (\mathcal{G}_c^{i-2} - 8\mathcal{G}_c^{i-1} + 8\mathcal{G}_c^{i+1} - \mathcal{G}_c^{i+2}) \quad (2.36)$$

and it can be shown that:

$$G_{c(4)}^{i+1/2} = \frac{7}{12} (g_c^{i+1} + g_c^i) - \frac{1}{12} (g_c^{i+2} + g_c^{i-1}). \quad (2.37)$$

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:

$$g_{p(4)}^{i+1/2} = \frac{G_{c(4)}^{i+1/2}}{\mathcal{V}_{c(4)}^{i+1/2}} \quad (2.38)$$

where $\mathcal{V}_{c(4)}$ is the fourth-order interpolated cell volume at $i + 1/2$.

2.3.4.2 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 [53] which we now describe.

Consider a density structure being advected at a constant velocity V in one dimension as shown in Fig. 2.5. 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

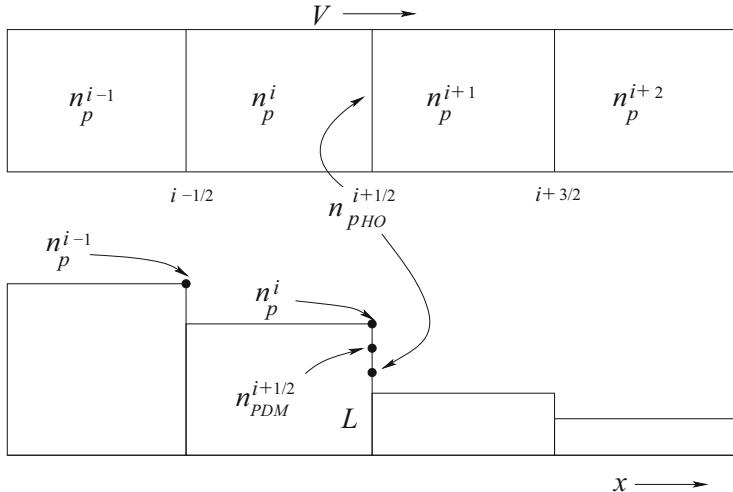


Fig. 2.5 Schematic of using the partial donor cell to determine the left state L of the density n

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. 2.5, 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.

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. 2.6. 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. 2.6. 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.:

$$(n_p^{i-1} - n_{PDM}^{i+1/2})V\Delta t = (n_p^{i-1} - n_p^i)\Delta x . \quad (2.39)$$

From (2.39), we find that:

$$n_{PDM}^{i+1/2} = \frac{1}{\alpha} [n_p^i - (\alpha - 1)n_p^{i-1}] \quad (2.40)$$

where $\alpha = V\Delta t/\Delta x$ is referred to as the partial donor method (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

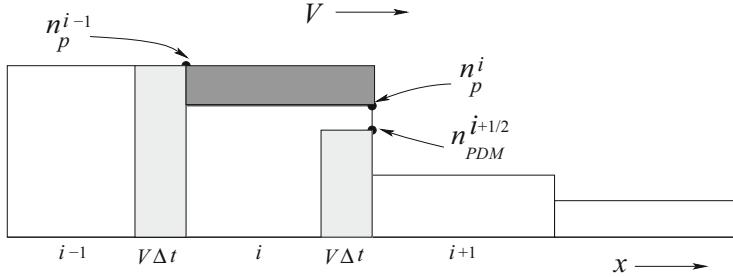


Fig. 2.6 Schematic showing how the partial donor cell value n_{PDM} is obtained

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.

Referring to Fig. 2.5, 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. 2.5, 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:

$$n_{PDM}^{i+1/2} = \frac{1}{\alpha} \left[n_p^i - (\alpha - 1)n_p^{i+1} \right]. \quad (2.41)$$

2.3.5 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 Ref. [51].

We define the following generalized transport equation:

$$\frac{\partial \rho \langle \chi \rangle}{\partial t} + \nabla \cdot \rho \langle \chi \mathbf{v} \rangle - \rho \left\langle \mathbf{a} \cdot \frac{\partial \chi}{\partial \mathbf{v}} \right\rangle = 0 \quad (2.42)$$

where:

$$\langle \mathcal{F} \rangle = \int d\mathbf{v} f(\mathbf{v}) \mathcal{F}$$

$$\mathbf{a} = \frac{\nabla \mathbf{M}}{8\pi\rho}; \quad \mathbf{M} = B_x^2 \mathbf{e}_x + B_y^2 \mathbf{e}_y + B_z^2 \mathbf{e}_z.$$

The continuity equation is obtained from (2.42) 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 (2.42), the 3D MHD equations (2.3)–(2.5) are recovered?

The following distribution function satisfies this requirement:

$$f = f_1 + f_2 \quad (2.43)$$

where

$$f_1 = \frac{\exp(-u_1^2)}{(\pi v_{1x}^2)^{1/2}} \frac{\exp(-v_1^2)}{(\pi v_{1y}^2)^{1/2}} \frac{\exp(-w_1^2)}{(\pi v_{1z}^2)^{1/2}} \quad (2.44)$$

$$f_2 = -(u_2 v_2 + u_2 w_2 + v_2 w_2) \frac{\exp(-u_2^2)}{(\pi v_{2x}^2)^{1/2}} \frac{\exp(-v_2^2)}{(\pi v_{2y}^2)^{1/2}} \frac{\exp(-w_2^2)}{(\pi v_{2z}^2)^{1/2}} \quad (2.45)$$

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 (2.43), 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 are calculated directly, i.e., one does not have to solve the hydrodynamic and magnetic force components separately or differently.

2.3.6 Magnetic Field Evolution

The evolution of the magnetic field is governed by:

$$\frac{\partial \mathbf{B}}{\partial t} = -c \nabla \times \mathbf{E} = -c \nabla \times [\mathbf{E}_c + \mathbf{E}_H] \quad (2.46)$$

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 (2.46) is as follows: The normal component of the field is integrated over a cell face:

$$\int d\mathbf{A} \cdot \frac{\partial \mathbf{B}}{\partial t} = -c \int d\mathbf{A} \cdot \nabla \times \mathbf{E} = -c \int \mathbf{E} \cdot d\mathbf{l} \quad (2.47)$$

where $\int d\mathbf{l}$ denotes a line integral around the cell face (see Fig. 2.1). The advantage of using this technique to calculate the magnetic field is that it 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:

$$\frac{\partial \Phi_x}{\partial t} = -c E_y dy - c E_z dz \quad (2.48)$$

where $\Phi_x = B_x dy dz$. The updated magnetic flux in the x direction is:

$$\Phi_x^{t+\Delta t} = \Phi_x^t - c \Delta t (E_y dy + E_z dz)^{t+\Delta t/2}. \quad (2.49)$$

The convective and Hall electric fields are structurally very different and require different numerical techniques. We describe each field calculation separately.

2.3.6.1 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:

$$E_{cz} = -\frac{1}{c} (V_x B_y - V_y B_x) \quad (2.50)$$

which is written as:

$$E_{cz} = -\frac{1}{c} \left(\int f(\mathbf{v}) v_x B_y d^3 v - \int f(\mathbf{v}) v_y B_x d^3 v \right). \quad (2.51)$$

Thus, the calculation of \mathbf{E}_c is based on the fact that $\mathbf{V} = \int f(\mathbf{v}) \mathbf{v} d^3 v$. We note that the actual implementation of this technique is fairly complicated; it is described in [51].

2.3.6.2 Hall Electric Field

The Hall electric field is:

$$\mathbf{E}_H = -\frac{1}{c} \mathbf{V}_H \times \mathbf{B} = \frac{1}{ne c} \mathbf{J} \times \mathbf{B}. \quad (2.52)$$

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:

$$E_{Hz} = -\frac{1}{c}(V_{Hx}B_y - V_{Hy}B_x) \quad (2.53)$$

which we break up into two parts:

$$E_{Hz}^{xy} = -V_{Hx}B_y \quad \text{and} \quad E_{Hz}^{yx} = V_{Hy}B_x. \quad (2.54)$$

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 (2.54).

We now focus on the component E_{Hz}^{xy} which is shown in Fig. 2.7. In this figure, E_{Hz}^{xy} and V_{Hx} are defined along a cell edge (as shown in Fig. 2.1) 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_{Hz}^{xy} is:

$$E_{Hz}^{xy} = \begin{cases} V_{Hx}B_{yL} & \text{for } V_{Hx} > 0 \\ V_{Hx}B_{yR} & \text{for } V_{Hx} < 0 \end{cases} \quad (2.55)$$

The interpretation of (2.55) is straightforward. If $V_{Hx} > 0$, as shown in Fig. 2.7, 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.

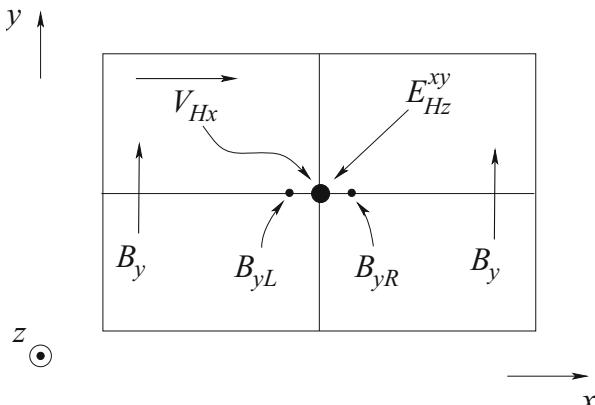


Fig. 2.7 Schematic of the upwinding scheme used to calculate one component of the Hall electric field

2.3.7 Courant Condition

The time step in an explicit code is controlled by a Courant condition:

$$\Delta t < \frac{\Delta x}{V_{max}} \quad (2.56)$$

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 wave number. We set $k = 2\pi/\Delta x$ to evaluate V_w numerically. Thus, we use the following velocity in the Courant condition:

$$V_{max} = \max(V + V_m + V_H + V_{hdw} + V_w). \quad (2.57)$$

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, $V_w \gg V_A$ 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 three-point smoothing algorithm [54]. 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.

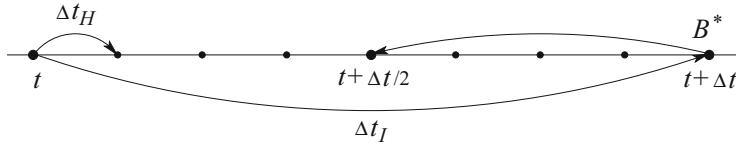


Fig. 2.8 Schematic showing the sub-cycling of the Hall term

2.3.8 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.

One sub-cycling time advance scheme is shown in Fig. 2.8. 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 (2.57). 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 Eqs. (2.3)–(2.6) on the Hall time scale. We only solve (2.52) on the Hall time scale. We advance the magnetic field using (2.52) 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 (2.3)–(2.5) are solved along with the convective magnetic induction equation:

$$\frac{\partial \mathbf{B}}{\partial t} = -c \nabla \times \mathbf{E}_c = \nabla \times \mathbf{V} \times \mathbf{B}. \quad (2.58)$$

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).

2.3.9 Implicit Numerical Scheme

As indicated above, explicit time schemes used to solve the Hall MHD equations are computationally expensive because the Courant condition is based on the whistler wave phase velocity. However, implicit time schemes are often used to overcome this problem. Chacón and Knoll [55] formulated an implicit scheme for incompressible Hall MHD in two dimensions. Subsequently, Tóth et al. [56] developed both an explicit and an implicit methodology to solve the Hall MHD equations on block-adaptive grids in 3D. They used the Newton-Krylov-Schwarz (NKZ) method in the implicit time integration of the Hall MHD equations. A number of test cases were presented: whistler wave propagation, magnetic reconnection, and magnetosphere modeling.

2.4 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.

2.4.1 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.

2.4.1.1 Whistler Waves

The linear wave relation for whistler waves is given by (2.15). 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 $^{-3}$ and $\beta = 10^{-4}$, the plasma is assumed to be isothermal (i.e., $T = \text{cnst}$), 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. 2.9, 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 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.

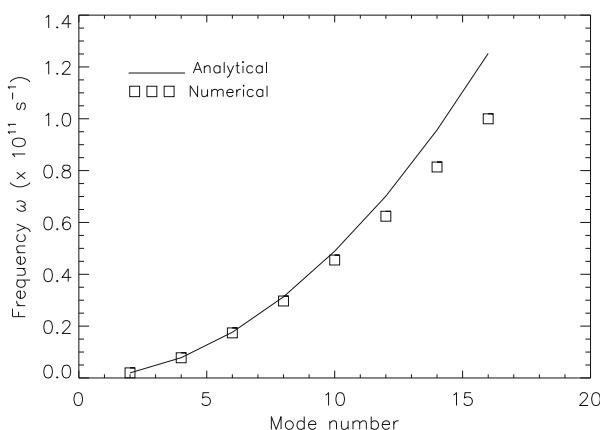


Fig. 2.9 Analytical and numerical comparison of the whistler wave frequency

2.4.1.2 Hall Drift Waves

A simple linear wave relation for Hall drift waves is given by (2.18) that is valid in the limit $L_n << c/\omega_{pi}$. However, the more general dispersion equation is given by Huba [47]:

$$\omega^2 - k_y^2 V_A^2 - \omega k_y V_A^2 / \Omega_{ci} L_n = 0 \quad (2.59)$$

for $\beta << 1$. The solution to this equation is:

$$\omega = \frac{k_y V_A^2}{2 L_n \Omega_{ci}} \pm \frac{1}{2} \left[\frac{k_y^2 V_A^4}{L_n^2 \Omega_{ci}^2} + 4 k_y^2 V_A^2 \right]^{1/2} \quad (2.60)$$

which illustrates the coupling between the Alfvén wave and the Hall drift wave. In the limit $L_n >> V_A / \Omega_{ci}$ (or $L_n >> c/\omega_{pi}$), (2.60) reduces to $\omega = k_y V_A$; in the opposite limit, it reduces to (2.18). 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, and the density is inhomogeneous with a profile:

$$n(x) = \frac{n_0}{2} [(1 + A) + (1 - A) \tanh((x - x_0)/\Delta x)] \quad (2.61)$$

where A is referred to as the Atwood number, $\beta = 10^{-4}$, and the plasma is assumed to be isothermal (i.e., $T = \text{const}$). 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 (2.61) 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. 2.10, we plot the normalized phase velocity of the Hall drift wave and the normalized density as a function of x . The density and velocity are normalized to the density and the Alfvén velocity at $x = 0$. 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. 2.11, 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. 2.11, 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.

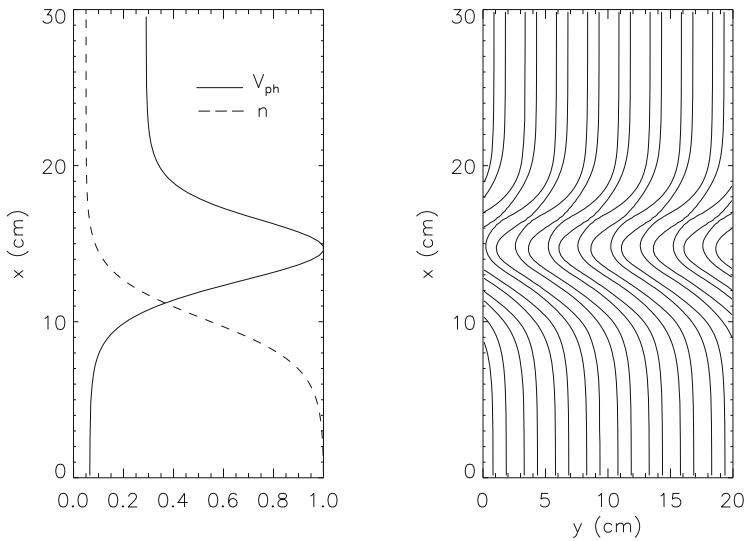
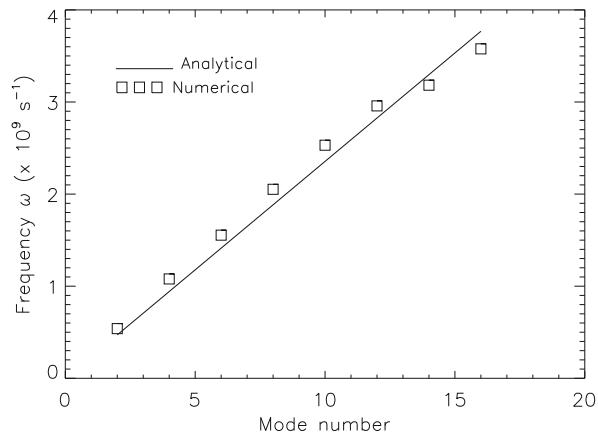


Fig. 2.10 Plot of the normalized phase velocity and density (left panel) and contour plot of the perturbed magnetic field (right panel) for the Hall drift wave

Fig. 2.11 Analytical and numerical comparison of the Hall drift wave frequency at $x = 15$ cm



2.4.2 Plasma Opening Switch

The plasma opening switch is a laboratory device designed to provide the appropriate power conditioning for inductive, pulsed, high-power generators in [9–11, 14, 15, 18]. A schematic of an opening switch is shown in Fig. 2.12. 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

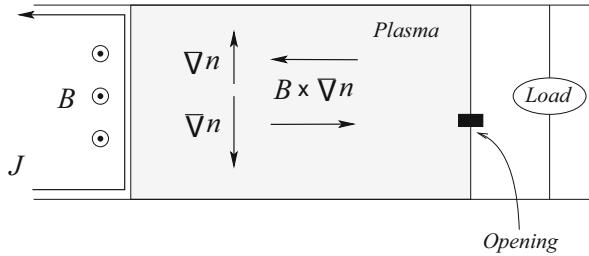


Fig. 2.12 Schematic of a plasma opening switch

transfers the stored magnetic energy to the load. The energy released is in the range 10–1000 KJ over a time scale 10s 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 $n > 10^{16} \text{ cm}^{-3}$, the plasma is controlled by ideal MHD (i.e., Alfvén time scales), while at low densities $n < 10^{13} \text{ cm}^{-3}$, 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 $L_x = 8 \text{ cm}$ and $L_y = 8 \text{ cm}$, and the mesh size is $(x, y) = (40, 50)$; the mesh is nonuniform in the y -direction. The density profile is:

$$n(y) = \frac{n_0}{Z} [(1 + \alpha \tanh(y/\Delta y))] \quad (2.62)$$

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 ($\mathbf{B} = B_0 \tanh(t/\tau_r) \hat{\mathbf{e}}_z$) is imposed at the $x = 0$ boundary with $B_0 = 1.2 \times 10^4 \text{ G}$ and $\tau = 10^{-6} \text{ s}$; this models the current rise time associated with the capacitive discharge. Finally, the plasma is assumed to be isothermal (i.e., $T = \text{const}$) and consists of doubly charged carbon ions ($Z = 2$ and $m_i = 12m_p$). We also assume $\beta = 10^{-4}$.

The results are shown in Fig. 2.13. 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 $n_0 = 10^{12} \text{ cm}^{-3}$. 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 $n_0 = 10^{12} \text{ cm}^{-3}$ at time $t = 20 \text{ ns}$. 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 ($\simeq 2 \mu\text{s}$). For $B_0 \simeq 200 \text{ G}$, we find

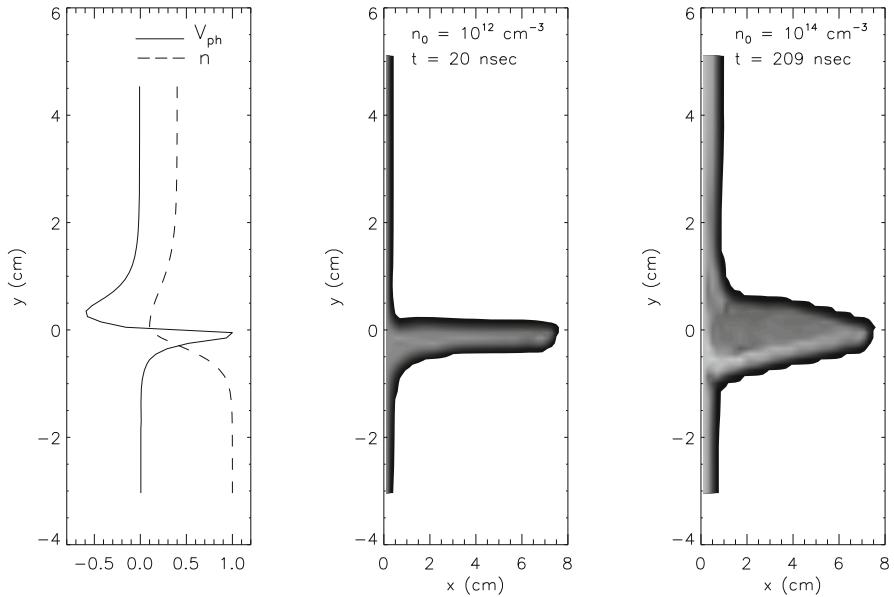


Fig. 2.13 Simulation of a plasma opening switch

that $V_A \simeq 6 \times 10^7 \text{ cm/s}$ and $V_{ph} \simeq 8 \times 10^9 \text{ cm/s}$. In the right panel, we show results for $n_0 = 10^{14} \text{ cm}^{-3}$. The Hall penetration time of the magnetic field is an order of magnitude longer than 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” its way into the plasma.

2.4.3 Sub-Alfvénic Plasma Expansions

Barium release experiments were carried out in the Earth’s magnetotail during the NASA AMPTE mission [1]. These releases had sufficient kinetic energy to create a diamagnetic cavity ($B \sim 0$) as the barium ions expanded and “swept up” the magnetic field. However, the expansion velocity of the barium ions (V_0) was sub-Alfvénic (V_A) where $V_0 \sim 1.3 \text{ km/s}$ and $V_A \sim 200 \text{ km/s}$. One of the intriguing and unexpected observations was that the barium cloud structured during the expansion phase of the release. Field-aligned striations formed on time scales less than an ion cyclotron period and on length scales less than an ion gyroradius. Hassam and Huba [2] proposed a new instability based on Hall MHD to explain this phenomena. The instability is an interchange mode similar to the Rayleigh-Taylor instability but has a faster growth rate and shorter wavelength, consistent with observations. This research spawned several laboratory experiments to explore the dynamics of sub-Alfvénic plasma expansions [3,5,7] as well as plasma simulation studies using Hall

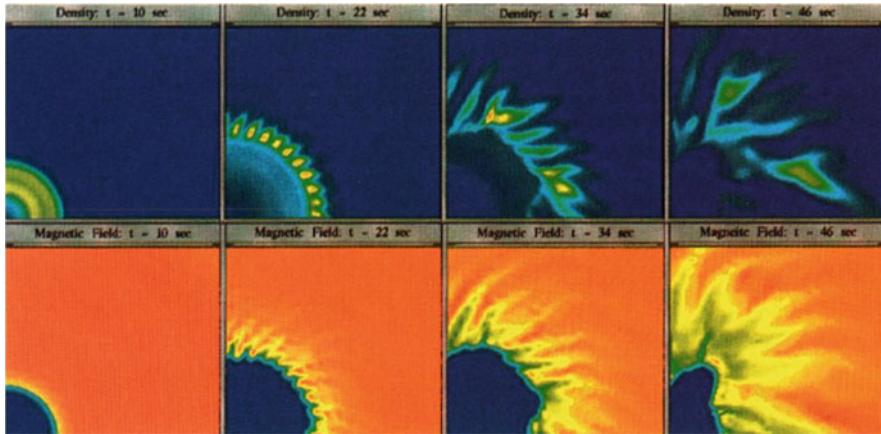


Fig. 2.14 Contours of the barium density (top panel) and magnetic field (bottom panel) as a function of time during a sub-Alfvénic expansion. Figure from [6] reprinted with permissions by the American Geophysical Union

MHD models [4, 6] and hybrid codes [57–60]. A recent, in-depth, review of this subject is given in Winske et al. [8]. As an example of the early 2D simulation results, we present Fig. 2.14 [6]. In this figure, we show the color contours of constant density (top panel) and magnetic field (bottom panel). The magnetic field is $\mathbf{B} = B_0 \hat{e}_z$ and orthogonal to the plane depicted in the figure. For the density, blue corresponds to the ambient value ($n_0 = 4.4 \times 10^2 \text{ cm}^{-3}$) and red to the maximum density ($(n_{\max} = 2.2 \times 10^5 \text{ cm}^{-3}$). For the magnetic field, blue corresponds to essentially no magnetic field ($B = 10^{-6} \text{ G}$) and red to $B = 1.50 \times 10^{-3} \text{ G}$ which is slightly larger than the ambient magnetic field ($B = 1.24 \times 10^{-3} \text{ G}$). The salient points are that the plasma rapidly structures during the expansion phase of the release and creates a diamagnetic cavity. At the end of the simulation, the diamagnetic cavity collapses as the background magnetic field penetrates the low-density cavities caused by the instability. However, structured plasma enhancements coalesce and continue to propagate outside the diamagnetic cavity.

We present 3D ideal and Hall MHD simulations of an expanding ion cloud in Figs. 2.15 and 2.16 [8]. For these simulations, the ions are deposited using:

$$\frac{\partial n_i}{\partial t} = n_0 \sigma_i (1 - \exp(-\sigma_i t)) \exp(-\sigma_i t) \exp(-((r - V_0 t)/\Delta r_0)^2) \quad (2.63)$$

where $n_0 = 5 \times 10^4 \text{ cm}^{-3}$ is the initial neutral density, $\sigma_i = 0.14 \text{ s}^{-1}$ is the ionization rate, $V_0 = 1 \text{ km/s}$ is the expansion velocity, and $\Delta r_0 = 0.5 \text{ km}$. The simulation is initialized at $t = 2 \text{ s}$ with an initial spherical shell of expanding ions at a radius $r = 2 \text{ km}$. Additionally, there is a 1D background magnetic field $B_z = 10^{-4} \text{ G}$. These parameters are chosen for numerical expediency and do not represent a specific magnetospheric barium release. The grid used is $100 \times 100 \times 100$ and is

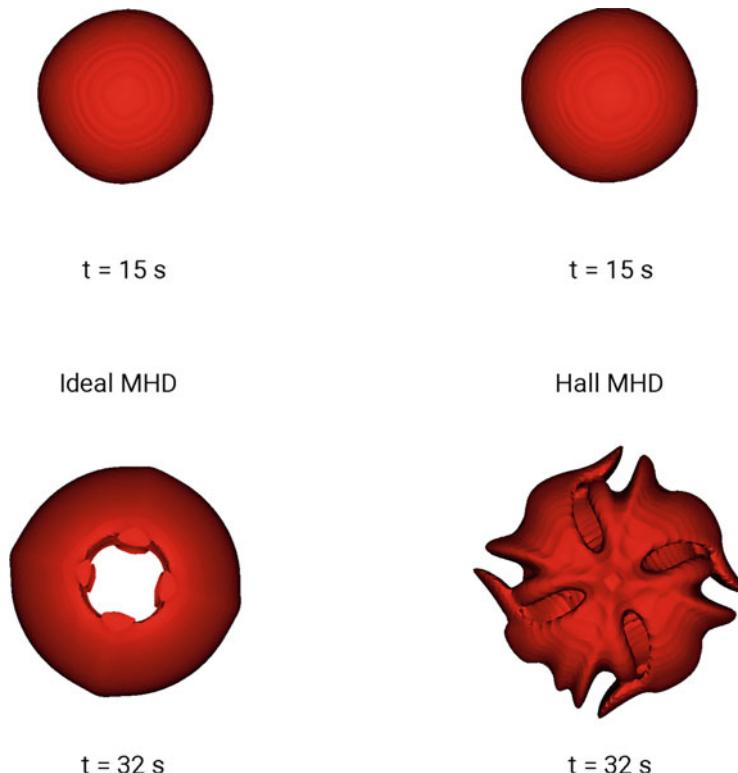


Fig. 2.15 Isodensity contours of the expanding ion cloud viewed along the magnetic field at times $t = 15$ s and 32 s for the ideal and Hall MHD cases

nonuniform; the extent of the grid perpendicular to B is $\pm 55 \lambda_i$ and along B is $\pm 140 \lambda_i$ where λ_i is the ion inertial length.

The results are shown in Figs. 2.15 and 2.16. We show an isodensity contour $n_i = 700 \text{ cm}^{-3}$ at times $t = 15$ s and 32 s for the ideal MHD case (left panel) and Hall MHD case (right panel). The z -axis is aligned with the background magnetic field. In Fig. 2.15, the view of the contours is along the z -axis, i.e., along the magnetic field. At $t = 15$ s, both the ideal and Hall MHD cases are essentially identical. However, at $t = 32$ s, the ideal MHD case is basically unstructured, while the Hall MHD case has become unstable, and large-scale density irregularities have developed. In Fig. 2.16, the view of the contours is perpendicular to the background magnetic field. Again, at $t = 15$ s, both the ideal and Hall MHD cases are essentially identical, but the contours are more extended along the magnetic field because the ions can freely expand along this axis. At $t = 32$ s, the ideal MHD case is not structured, and the extension along the magnetic field is more pronounced. The Hall MHD case shows the development of density irregularities that are filamented along the magnetic field, as expected.

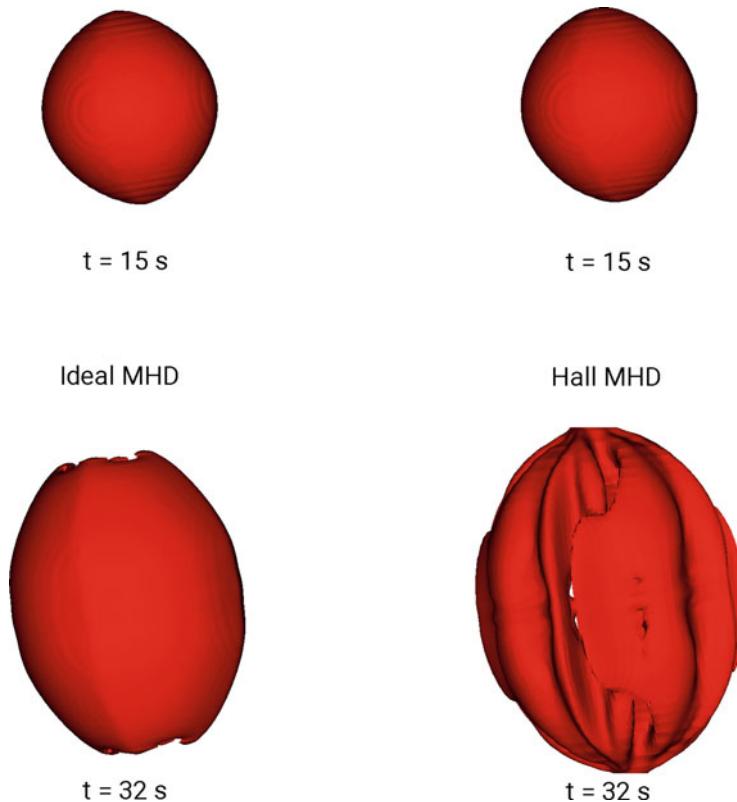


Fig. 2.16 Isodensity contours of the expanding ion cloud viewed perpendicular to the magnetic field at times $t = 15$ and 32 s for the ideal and Hall MHD cases

2.4.4 Magnetic Reconnection

In recent years, it has become evident that Hall physics plays a critical role in magnetic reconnection processes [19–32]. 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 [21, 22, 24, 29] 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 $c/\omega_{pi} \sim 10\text{s}-100\text{s km}$. 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.

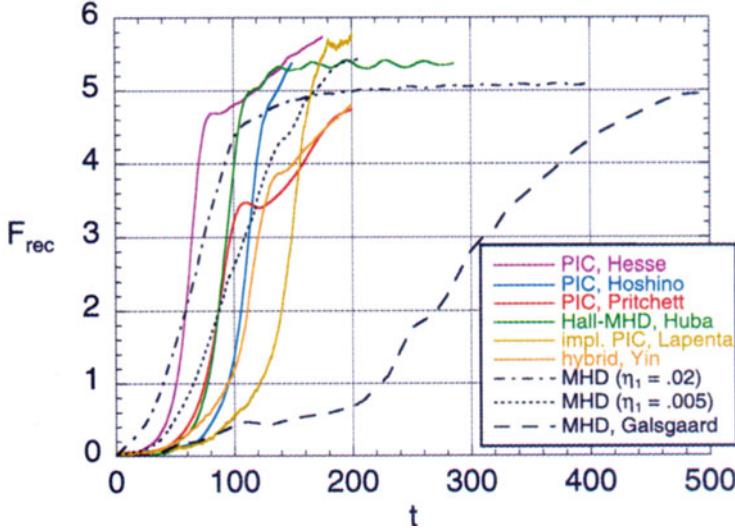


Fig. 2.17 Comparison of the magnetic reconnection rate from different models. Figure from [37] reprinted with permissions by the American Geophysical Union

Seminal research papers on the subject are the Geospace Environment Modeling (GEM) magnetic reconnection challenge [25] and the “Newton Challenge” forced magnetic reconnection project [37]. These challenges involved comparing the magnetic reconnection rate obtained from different 2D physics-based models: particle-in-cell (PIC), hybrid, Hall MHD, and ideal MHD. The basic result is shown in Fig. 2.17 which plots the magnetic reconnection flux (F_{rec}) versus time (t). Although there are different times for the onset of reconnection, the PIC, hybrid, and Hall MHD model results have essentially the same reconnection rate, i.e., the slope of the curves between, say, $F_{rec} = 1 - 4$. The ideal MHD reconnection rate is much slower (dashed black curve) but can be increased by imposing an enhanced resistivity (η_i).

Subsequent work by Huba and Rudakov [40] found that the asymptotic (i.e., time independent) state of the system is nearly independent of the initial current sheet width. Specifically, the Hall reconnection rate is weakly dependent on the initial current layer width and is $\partial\Phi/\partial t \lesssim 0.1 V_{A0} B_0$ where Φ is the reconnected flux, and V_{A0} and B_0 are the Alfvén velocity and magnetic field strength in the upstream region. Moreover, this rate appears to be independent of the scale length on which the electron “frozen-in” condition is broken (as long as it is $< c/\omega_{pi}$) and the system size. This is shown in Fig. 2.18 where we plot gray-scale contours of the plasma density for the initial and final states of the system as a function of x and y for different values of the initial current sheet width L : $L = 1, 5, 10, 20$. The scale length L is normalized to the ion inertial length c/ω_{pi0} based on the density n at $x = 0$. The density of the initial states is normalized to $n_0/2$, and the density of the final states is normalized to $n_0/3$. The white contour denotes high density, and

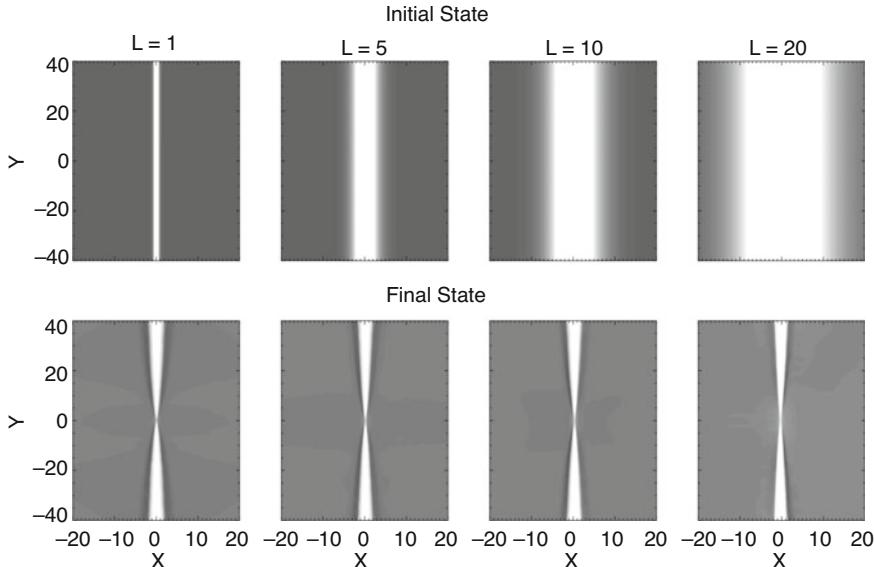


Fig. 2.18 Gray-scale contours of the plasma density for the initial and final states of the system as a function of x and y for different values of the initial current sheet width: $L = 1, 5, 10$, and 20 . Figure from [40] reprinted with permissions by the American Physical Society

the dark contour denotes low density. The important result is that the final states for all four cases are very similar despite the very different initial conditions. This strongly suggests that the asymptotic state of a Hall-dominated reconnection plasma is independent of initial conditions: specifically, the width of 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 . Physically, this component of the wave field is associated with the “dragging” of the field out of the reconnection plane by electrons within the current layer because they are $E \times B$ drifting, while the ions are essentially stationary because they are unmagnetized. Additionally, another difference is the asymmetric propagation of the reconnection X-line in Hall MHD [38].

We illustrate these new features of Hall MHD reconnection in Fig. 2.19 using VooDoo [51]. The 2D planes are in the x/y direction where x is “horizontal” and y is “vertical.” The white contour lines at $z = 0$ are the magnetic stream lines where the magnetic field is $B_x = B_0 \tanh(y/L)$. The colored x/y planes depict the plasma density showing the current layer at $y = 0$. The red/yellow isosurfaces are of B_y . The top panel is at $t = 0$ where the system is perturbed at $z \sim -20$. The bottom panel is after the reconnection process has proceeded for some time. The two key points are as follows: (1) the reconnected magnetic field stream lines at $z = 0$ are “bent” in the $-z$ direction which is the signature of the quadrupole Hall-generated field and (2) the X-line is propagating in the $-z$ direction. The propagation of the

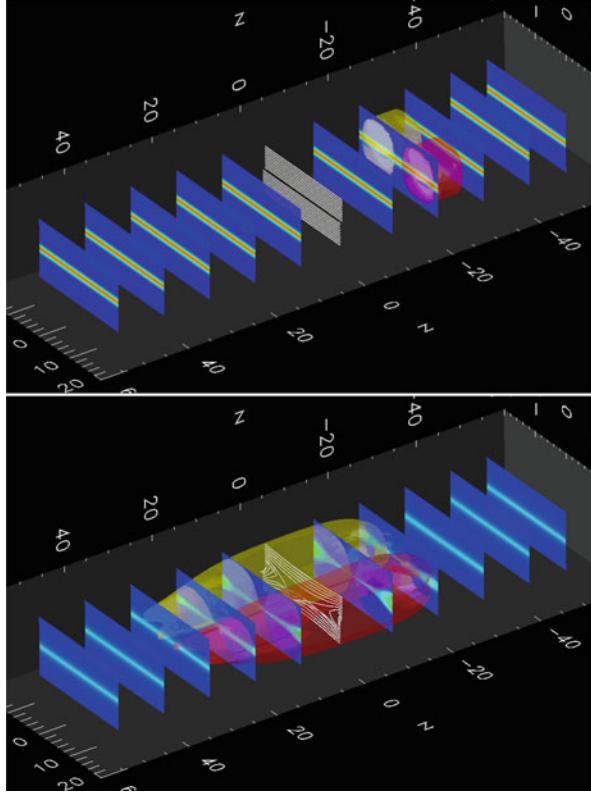


Fig. 2.19 3D simulation of magnetic reconnection at $t = 0$ (top panel) and at a later time (bottom panel). Details are described in the text

X-point in the direction of the current was also found in implicit particle-in-cell simulations of 3D magnetic reconnection [61].

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. (2.6)] in the electron magnetohydrodynamic limit ($\mathbf{V} = 0$) as:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times \left[\frac{c}{4\pi ne} \left(\nabla B^2/2 - (\mathbf{B} \cdot \nabla)\mathbf{B} \right) \right]. \quad (2.64)$$

When the magnetic field is unidirectional, i.e., no magnetic curvature, only the first term in Eq. (2.64) 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 [50]. However, in a reconnection geometry near the neutral sheet, magnetic field line curvature dominates, and the second term in Eq. (2.64)

controls the evolution of the magnetic field in the Hall limit. The evolution of B_y is approximately described by:

$$\frac{\partial B_y}{\partial t} \simeq -\frac{\partial}{\partial z} \frac{cB_y}{4\pi ne} \frac{\partial B_x}{\partial y} \simeq \frac{J_z}{ne} \frac{\partial B_y}{\partial z}. \quad (2.65)$$

We can rewrite Eq. (2.65) 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 [10]). Using (2.65), 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. (2.65) underestimates the wave speed because ion motion is neglected.

We now discuss the inclusion of a guide field as it pertains to Hall magnetic reconnection. There have been several studies that included a guide field and found that the reconnection rate is reduced [31, 35, 62, 63] and that the quadrupole field associated with Hall reconnection disappears [35, 64]. Pritchett and Coroniti [63] used a three-dimensional particle-in-cell code to investigate magnetic reconnection in the presence of a guide field. They found for moderate guide fields $B_{gf} \lesssim B_0$ that the reconnection rate was comparable to the zero guide field case; for stronger guide fields $B_{gf}/B_0 = 3$ and 5, the reconnection rate was roughly 40% of the $B_{gf}/B_0 = 1$ value. However, these simulations were not run until a steady state was achieved. A 2D simulation study was conducted [65] in which a steady state is achieved for $B_{gf}/B_0 = 0, 0.1, 0.2, 0.5, 1.0, 2.0$, and 5.0 where B_{gf} is the guide field and B_0 is the reversed field. It was found that the reconnection rate and plasma energization are reduced for increasing guide field strength. This is caused by a $\mathbf{J} \times \mathbf{B}$ force associated with Hall currents and the guide field that reduce the inflow and outflow velocities. However, the reconnection rate and plasma energization are only reduced by a factor of 2 for $B_{gf} = 5 B_0$. Additionally, the quadrupole field associated with Hall reconnection is eliminated for $B_{gf} \simeq B_0/3$.

We illustrate the 3D topological changes in the magnetic field and current layer for Hall magnetic reconnection including a guide field in Fig. 2.20. The simulation is similar to that shown in Fig. 2.19 except that a guide field is included with $B_z = B_0$. Again, the top panel is at $t = 0$, and the bottom panel is after the reconnection process has proceeded for some time. The white contour lines are magnetic field streamlines that are initially above the current layer at $y = 0$ and the red contour lines below the current layer. As in the case of no guide field, the X-line is propagating in the $-z$ direction. However, we find that reconnection process causes the magnetic field to form a helical topology leading to the formation of O-points in the center of the current layer and X-points are either side.

In recent years, there has been substantially more research on magnetic reconnection using Hall MHD models as well as sophisticated hybrid and particle-in-cell

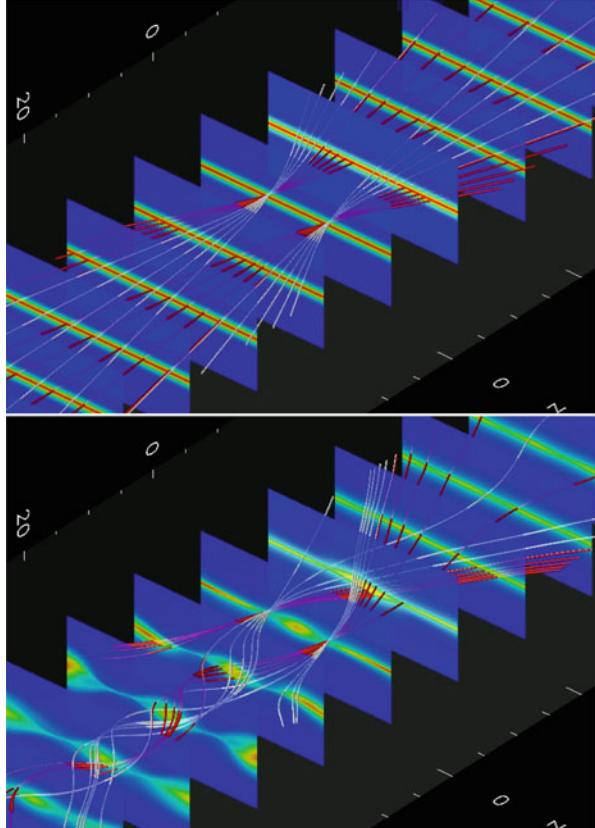


Fig. 2.20 3D simulation of magnetic reconnection at $t = 0$ (top panel) and at a later time (bottom panel) with a guide field. Details are described in the text

(PIC) models [41–44]. It has been shown in models using full ion physics that there are shortcomings in the Hall MHD description presented in this section. To a certain extent, this is not surprising since a simple isothermal pressure model is used and neglects ion pressure anisotropies associated with finite Larmor radius (FLR) effects and ion kinetic effects. In the absence of a guide field, [41] found that the peak Hall MHD reconnection rate overestimates the peak rate from hybrid and PIC models by $\sim 85\%$. We note that the Hall MHD reconnection rates presented in [41] are also much larger than those reported in [40]; the differences may be due to the specific simulation parameters or perhaps the numerical algorithms used. With regard to systems with a guide field, noted in [42] “the Hall-MHD fluid model was found to reproduce kinetic reconnection rates only for a sufficiently strong guide field, for which ion inertia breaks the frozen-in condition and the outflow becomes Alfvénic in the kinetic system.” Additionally, there were significant differences in multiple magnetic island coalescence between Hall MHD and kinetic simulations.

2.5 Summary

We have presented a discussion of 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. Three applications of Hall MHD physics are presented: (1) the plasma opening switch which illustrates rapid magnetic field transport associated with Hall drift wave, (2) the structuring of sub-Alfvénic plasma expansions, and (3) 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.

One application of Hall MHD physics to space physics is the global modeling of magnetospheres. One problem faced in this effort is the size of the ion inertial length relative to the size of the magnetosphere. In the Earth’s magnetosphere, the ion inertial length is ~ 100 km, while the size of the magnetosphere is $\sim 5 \times 10^6$ km. It is not feasible to model the Earth’s magnetosphere on grid sizes that resolve the ion inertial length at this time because of computational limitations. One strategy to address the problem is to model Hall MHD physics in limited regions of the magnetosphere where it is believed to be important, e.g., at the subsolar magnetopause and in the magnetotail. This method has been incorporated in the BATS-R-US code [66] in which Hall MHD physics provides improved boundary conditions on a sub-gridded PIC domain.

Another interesting application of global Hall MHD modeling is to the Ganymede magnetosphere which is much smaller than the Earth’s magnetosphere. The aforementioned BATS-R-US global Hall MHD model was used to study the solar wind interaction with Ganymede’s magnetosphere [67]. It was found Hall currents generated within the magnetopause and magnetotail led to a new set of field-aligned currents and a global system of ion drift belts that circulate Jovian magnetosphere plasma in the Ganymede magnetosphere.

Lastly, a global Hall MHD simulation of the lunar mini-magnetosphere has been reported [68]. In this study, the model was used to study the interaction of the solar wind with a local lunar crustal field. It was found that solar wind ions penetrated the magnetopause leading to the merging of the shock and magnetopause, as well as the development of plasma and magnetic field asymmetries.

Acknowledgments I am indebted to Dr. J. Lyon who developed a 2D Hall MHD code at NRL in the mid-1980s; 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 (deceased), Leonid Rudakov, Clark Mobarry, Joel Fedder, Dan Winske, and Adil Hassam.

References

1. P.A. Bernhardt, R.A. Roussel-Dupré, M.B. Pongratz, G. Haerendel, A. Valenzuela, D. Gurnett, R.R. Anderson, *J. Geophys. Res.* **92**, 5777 (1987)
2. A.B. Hassam, J.D. Huba, *Geophys. Res. Lett.* **14**, 60 (1987)
3. B.H. Ripin, E.A. McLean, C.K. Manka, C. Pawley, J.A. Stamper, T.A. Peyser, A.N. Mostovych, J. Grun, A.B. Hassam, J.D. Huba, *Phys. Rev. Lett.* **59**, 2299 (1987)
4. J.D. Huba, J.G. Lyon, A.B. Hassam, *Phys. Rev. Lett.* **59**, 2971 (1987)
5. G. Dimonte, L.G. Wiley, *Phys. Rev. Lett.* **67**, 1755 (1991)
6. J.D. Huba, P.A. Bernhardt, J.G. Lyon, *J. Geophys. Res.* **97**, 11 (1992)
7. B.H. Ripin, J.D. Huba, E.A. McLean, C.K. Manka, T.A. Peyser, H.R. Burris, J. Grun, *Phys. Fluids B* **5**, 3491 (1993)
8. D. Winske, J.D. Huba, C. Niemann, L. Ari, Recalling and updating research on diamagnetic cavities: Experiments, Theory, Simulations. *Frontiers in Astr. and Space Sci.* **5**, 51 (2019). <https://doi.org/10.3389/fspas.2018.00051>
9. R.A. Meger, R.J. Comisso, G. Cooperstein, S.A. Goldstein, *Appl. Phys. Lett.* **42**, 943 (1983)
10. B.V. Weber, R.J. Comisso, R.A. Meger, J.M. Neri, W.F. Oliphant, P.F. Ottlinger, *Appl. Phys. Lett.* **45**, 1043 (1984)
11. G. Cooperstein, P.F. Ottlinger, Guest Editorial, *IEEE Trans. Plasma Sci.* **PS-15**, 629 (1987)
12. A.S. Kingsep, K.V. Chukbar, V.V. Yan'kov, *Reviews of Plasma Physics*, vol. 16, ed. by B.B. Kadomtsev (Consultants Bureau, New York, 1990)
13. A. Fruchtman, *Phys. Fluids B* **3**, 1908 (1991)
14. R.J. Comisso, P.J. Goodrich, J.M. Grossmann, D.D. Hinshelwood, P.F. Ottlinger, B.V. Weber, *Phys. Fluids B* **4**, 2368 (1992)
15. D.D. Hinshelwood, B.V. Weber, J.M. Grossmann, R.J. Comisso, *Phys. Rev. Lett.* **68**, 3567 (1992)
16. R.J. Mason, P.L. Auer, R.N. Sudan, B.V. Oliver, C.E. Seyler, J.B. Greenly, *Phys. Fluids B* **5**, 1115 (1993)
17. J.D. Huba, J.M. Grossmann, P.F. Ottlinger, *Phys. Plasmas* **1**, 3444 (1994)
18. A. S. Chuvatin, B. Etlicher, *Phys. Rev. Lett.* **74**, 2965 (1995)
19. J.F. Drake, R.G. Kleva, M.E. Mandt, *Phys. Rev. Lett.* **73**, 1251 (1994)
20. D. Biskamp, E. Schwarz, J.F. Drake, *Phys. Rev. Lett.* **75**, 3850 (1995)
21. M.A. Shay, J.F. Drake, R.E. Denton, D. Biskamp, *J. Geophys. Res.* **103**, 9165 (1998)
22. M.A. Shay, J.F. Drake, *Geophys. Res. Lett.* **25**, 3759 (1998)
23. M. Rastätter, M. Hesse, K. Schindler, *J. Geophys. Res.* **104**, 12301 (1999)
24. M.A. Shay, J.F. Drake, B.N. Rogers, R.E. Denton, *Geophys. Res. Lett.* **26**, 2163 (1999)
25. 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)
26. M. Hesse, J. Birn, M. Kuznetsova, *J. Geophys. Res.* **106**, 3721 (2001)
27. J. Birn, M. Hesse, *J. Geophys. Res.* **106**, 3737 (2001)
28. A. Otto, *J. Geophys. Res.* **106**, 3751 (2001)
29. M.A. Shay, J.F. Drake, B.N. Rogers, R.E. Denton, *J. Geophys. Res.* **106**, 3759 (2001)
30. Z.W. Ma, A. Bhattacharjee, *J. Geophys. Res.* **106**, 3773 (2001)
31. P. Pritchett, *J. Geophys. Res.* **106**, 3783 (2001)
32. L. Yin, D. Winske, S.P. Gary, J. Birn, *J. Geophys. Res.* **106**, 10761 (2001)
33. P.L. Pritchett, *J. Geophys. Res.* **106**, 25961 (2001)
34. M. Hesse, M. Kuznetsova, J. Birn, *J. Geophys. Res.* **106**, 29831 (2001)
35. H. Karamabadi, D. Krauss-Varban, N. Omidi, H.X. Vu, *J. Geophys. Res.* **104**, 12313 (1999)
36. Z.W. Ma, L.C. Lee, *J. Geophys. Res.* **106**, 25951 (2001)
37. J. Birn, K. Galsgaard, M. Hesse, M. Hoshino, J.D. Huba, G. Lapenta, P.L. Pritchett, K. Schindler, L. Yin, J. Büchner, T. Neukirch, E.R. Priest, *Geophys. Res. Lett.* **32**, L03102 (2005). <https://doi.org/10.1029/2004GL021491>
38. J.D. Huba, L.I. Rudakov, *Phys. Plasmas*, **9**, 4435 (2002)

-
39. J.D. Huba, L.I. Rudakov, Phys. Plasmas, **10**, 3139 (2003)
40. J.D. Huba, L.I. Rudakov, Phys. Rev. Lett. **93**, 175003 (2004)
41. A. Stanier, W. Daughton, L. Chacón, H. Karimabadi, J. Ng, Y.-M. Huang, A. Hakim, A. Bhattacharjee, Phys. Rev. Lett. **115**, 175004 (2015)
42. A. Stanier, W. Daughton, A.N. Simakov1, L. Chacón, A. Le1, H. Karimabadi, J. Ng, A. Bhattacharjee, Phys. Plasmas **24**, 022124 (2017) <https://doi.org/10.1063/1.4976712>
43. Y.-H. Liu, W. Daughton, H. Karimabadi, H. Li, S.P. Gary, Phys. Plasmas **21**, 022113 (2014). <https://doi.org/10.1063/1.4865579>
44. J. Ng, Y.-M. Huang, A. Hakim, A. Bhattacharjee, A. Stanier, W. Daughton, L. Wang, K. Germaschewski, Phys. Plasmas **22**, 112104 (2015). <https://doi.org/10.1063/1.4935302>
45. Huba, J.D., A Tutorial on Hall Magnetohydrodynamics, in *Space Simulations*, eds. by M. Scholer, C.T. Dum, J. Büchner (Springer, New York, 2003), p. 170
46. J.A. Bittencourt, *Fundamentals of Plasma Physics* (Pergamon Press, New York, 1986)
47. J.D. Huba, Phys. Fluids B **3**, 3217 (1991)
48. A. Fruchtman, L.I. Rudakov, Phys. Rev. Lett. **69**, 2070 (1992)
49. A. Fruchtman, L.I. Rudakov, Phys. Rev. E**50**, 2997 (1994)
50. L.I. Rudakov, J.D. Huba, Hall magnetic shocks in plasma current layers. Phys. Rev. Lett. **89**, 095002 (2002)
51. J.D. Huba, J.G. Lyon, J. Plasma Phys. **61**, 391 (1999)
52. S. Zalesak, J. Comp. Physics **31**, 335 (1979)
53. K. Hain, J. Comput. Phys. **73**, 131 (1987)
54. C.K. Birdsall, A.B. Langdon, *Plasma Physics via Computer Simulation* (Adam Hilger, New York 1991)
55. L. Chacón, D.A. Knoll, J. Comput. Phys. **188**, 573 (2003)
56. G. Tóth, Y. Ma, T.I. Gombosi, J. Comput. Phys. **227**, 6967 (2008)
57. D. Winske, J. Geophys. Res., **93**, 2539 (1988)
58. D. Winske, Phys. Fluids B **1**, 1900 (1989)
59. D. Winske, Phys. Plasmas **3**, 3966 (1996)
60. J.D. Huba, D. Winske, Phys. Plasmas **5**, 2305 (1998)
61. G. Lapenta, D. Krauss-Varban, H. Karimabadi, J.D. Huba, L.I. Rudakov, P. Ricci, Geophys. Res. Lett. **33**, L10102 (2006). <https://doi.org/10.1029/2005GL025124>
62. M. Hesse, M. Kuznetsova, M. Hoshino, Geophys. Res. Lett. **29**, 1563 (2002). <https://doi.org/10.1029/2001GL014714>
63. P.L. Pritchett, F.V. Coroniti, J. Geophys. Res. **109**, A01220 (2004). <https://doi.org/10.1029/2003JA009999>
64. B.N. Rogers, R.E. Denton, J.F. Drake, J. Geophys. Res. **108**, 1111 (2003). <https://doi.org/10.1029/2002JA009699>
65. J.D. Huba, Phys. Plasmas **12**, 012322 (2005)
66. L.K.S. Daldorff, G. Tóth, T.I. Gombosi, G. Lapenta, J. Amaya, S. Markidis, J.U. Brackbill, J. Comp. Phys. **268**, 236 (2014)
67. J.C. Dorelli, A. Glocer, G. Collinson, G. Tóth, J. Geophys. Res. Space Physics **120**, 5377 (2015). <https://doi.org/10.1002/2014JA020951>
68. L. Xie, L. Li, Y. Zhang, Y. Feng, X. Wang, A. Zhang, L. Kong, J. Geophys. Res. Space Phys. **120**, 6559 (2015) <https://doi.org/10.1002/2015JA021647>



Hybrid-Kinetic Approach: Massless Electrons

3

Dan Winske, Homa Karimabadi, Ari Yitzchak Le, Nojan Nick Omidi,
Vadim Roytershteyn, and Adam John Stanier

Abstract

Hybrid-kinetic approaches are widely used to model ion-scale phenomena in space plasmas. Hybrid codes differ from fully kinetic particle-in-cell (PIC) codes in that the electrons are modeled as a fluid that can be considered even massless as in this chapter, while the electric field is not advanced in time, but instead calculated at the new time level from the advanced ion quantities and the magnetic field. In this chapter we concentrate on such hybrid models with massless electrons, beginning with a discussion of the basics of a simple hybrid code algorithm. We then show examples of recent use of hybrid codes for large-scale space plasma simulations of structures formed at planetary bow shock—foreshock systems, magnetic reconnection at the magnetopause, and complex phenomena in the magnetosheath due to the interaction of kinetic processes associated with the bow shock, magnetic reconnection, and turbulence. A discussion then follows of a number of other hybrid codes based on different algorithms that are presently in active use to investigate a variety of plasma processes in space as well as some recent work on the development of new models.

D. Winske (✉) · A. Y. Le · A. J. Stanier
Los Alamos National Laboratory, Los Alamos, NM, USA
e-mail: winske@losalamos.com; arile@lanl.gov; stanier@lanl.gov

H. Karimabadi
Analytics Ventures, Inc., San Diego, CA, USA
e-mail: homa@alphatrail.com

N. N. Omidi
Solana Scientific Inc., Solana Beach, CA, USA
e-mail: omidi@roadrunner.com

V. Roytershteyn
Space Science Institute, Boulder, CO, USA
e-mail: vroytershteyn@spacescience.org

We conclude with a few brief comments concerning the future development and use of hybrid codes.

3.1 Introduction

A hybrid code in the context of plasma physics generally refers to a computational model in which some of the plasma species are treated kinetically and others as fluids, while the electric and magnetic fields can be considered as either electrostatic or electromagnetic. Most often in hybrid codes employed in space plasmas all the different ion species are treated kinetically using particle-in-cell methods, the electrons are modeled as a fluid with massless electrons, and Maxwell's equations are solved in the radiation-free (i.e., no light waves) limit. It is in this more restricted domain that we discuss hybrid codes in this chapter. Even in this limited space, there is still freedom in how to treat the fast phenomena associated with electron dynamics. In space plasmas in particular, hybrid codes occupy a unique niche, intermediate between fluid codes that are most useful for modeling large portions of the solar wind or the interaction of the solar wind with the Earth's magnetosphere and fully kinetic treatment of all the plasma components, i.e., all ion species and electrons, that are resolved down to the smallest electron scales. Fully kinetic codes are commonly referred to as PIC codes, because traditionally particle-in-cell methods have been used to treat the particle dynamics. Instead, hybrid codes model a portion of the solar wind-magnetosphere boundary or a small region in the solar wind or the magnetosphere on spatial and temporal scales on which ion dynamics dominates. The effects of the larger fluid scales can be modeled through boundary conditions, while the small-scale effects of electrons are added in through models for the electron pressure tensor, and high-frequency (e.g., electron-ion lower-hybrid or higher electron-electron) wave effects are included through a resistivity or other transport coefficients. But as computers have become larger and faster, fully kinetic codes can now model system sizes and time scales that previously were only accessible to hybrid codes, and hybrid codes in turn can now model systems large enough that could once only be studied by fluid models. Thus, as the domains of calculations with different physics models can overlap, having an assortment of hybrid models allows comparisons between hybrid-PIC and fluid-hybrid simulations to be done effectively in order to verify the validity of models, provide additional validation of codes, and give further insight into spacecraft data.

This chapter is an updated version of the Chapter "Hybrid Simulations Codes: Past, Present and Future—A Tutorial" in the 2003 edition of the book [1]. As we shall see, computational plasma physics has advanced so rapidly in recent years that what we called "The Future" in 2003 is now not even in the "The Present," but already fading into "The Past." However, some things have remained the same—namely, the basic algorithms used in hybrid codes. On the other hand, new advances in algorithm development show excellent promise for future major advances in

understanding of space physics phenomena. And present-day simulations have far exceeded what was thought to be possible 20 years ago.

In contemporary terms, the 2003 hybrid code tutorial had both strong and weak points. The old tutorial was very strong in its rather complete discussion of the assumptions and underlying equations and numerical methods of solution of the equations of the computational model as well as an extensive list of references. There was also a detailed historical tracing of the development of hybrid codes and associated numerical methods, particularly in relation to the 6th International Space Simulation School (ISSS-6) as well as to the earlier schools. The discussion included the most common implementations of the hybrid model in simulation codes as well as examples of results from calculations and comparative test cases. Most of the material in this tutorial is still relevant today and does not need to be repeated here. In addition, issues of great interest in the time frame of ISSS-6 were considered in some detail. For example, the question of how to compute the complete electron pressure tensor and ways to include it in a generalized hybrid model for application to magnetic reconnection were well described. Also, initial attempts at using hybrid codes to model global interactions of the solar wind with planetary magnetospheres were demonstrated.

However, tutorial [1] suffers from now being rather out of date concerning what is presently state-of-the-art. For example, the goals of performing efficient calculations on parallel computers with thousands of processors and analyzing data from multi-satellite missions, such as CLUSTER and THEMIS, were thought at the time to be perhaps a decade in the future. In fact, however, this was achieved much sooner and present-day simulations have extended far beyond that frontier. PIC and hybrid codes are now at the threshold of running in three spatial dimensions for millions of processor hours using trillions of particles, billions of cells, and almost a million time steps. What were then very difficult issues related to trying to embed hybrid or kinetic models in larger-scale fluid models in order to model micro-scale processes occurring in magnetic reconnection have now been replaced by very large-scale PIC codes that include all the smallest scales naturally in problems approaching global scales with realistic ratios of the ion to electron mass.

In addition to these very massive PIC and hybrid codes running on what is headed toward exascale computing, the computational environment is also changing in other ways. With larger computational domains, longer times, and more realistic initial/boundary conditions, the emphasis is now on investigating the interactions between various physical processes, rather than focusing on a single physics issue. This has been motivated in part by the existence of better spacecraft data, involving very high-resolution simultaneous data from multiple satellites. One example is the MMS (Magnetosphere Multi-Scale) mission [2] whose four closely-spaced satellites have fully resolved ion and have begun to explore electron scales related to magnetic reconnection layers. Better diagnostics in the simulations have led to much clearer comparisons and analysis of the spacecraft observations as well as improved insight into the underlying physical processes. With larger and longer duration calculations, the need for more sophisticated hybrid algorithms that pay careful attention to energy conservation and suppression of slow-growing numerical

instabilities has become a major issue, as we will discuss later. In addition, extended hybrid models that include electron inertia effects, new quasi-neutral models, and better Vlasov methods have become more prevalent, which allow investigating short-scale electron processes that are now becoming accessible in space and laboratory observations of magnetic reconnection. We touch only briefly on these topics later, which are treated in much more detail in Chaps. 5, 9, and 10.

This chapter on hybrid codes has several purposes. The first purpose is to serve as a tutorial for newcomers to the field of computational space plasma physics by providing a brief introduction to hybrid code methodology. It includes a review of the basics of the hybrid physics model, the underlying fundamental equations that are used, and a simple example of the physics implementation into a working algorithm. This Sect. 3.2 is relatively brief, but we emphasize in the discussion that there are old as well as recent articles that cover this background material quite completely and provide updated references. It also raises the major issue inherent in all hybrid algorithms, namely how to consistently update the electric field. This issue will also appear prominently later, in Sects. 3.4 and 3.5. The purpose of Sect. 3.3 is to provide some contemporary examples of hybrid code use in space physics and their relation to observations. This usage is illustrated in only a few examples, but appropriate references to these and other works are presented here as well as throughout the rest of the chapter. Section 3.4 addresses other hybrid algorithms and briefly describes their use for a wide range of applications at present. Section 3.5 discusses several new hybrid algorithm development efforts, the motivation for constructing them, and their potential applications as well as appropriate numerical test cases of their use. The inclusion of electron inertia effects in advanced hybrid algorithms is also briefly summarized. Finally, in Sect. 3.6 we look out into the future with some trepidation, recalling from the earlier version of this tutorial how rapidly and dynamically computational plasma physics can change, to suggest potential areas of growth, development, and use for hybrid codes in space and astrophysical plasmas.

3.2 Review of Basic Model and Implementation

Hybrid codes model plasma and low-frequency electromagnetic wave phenomena on spatial and temporal scales shorter than those used in a magnetohydrodynamic description, but longer than those that resolve electron dynamics (i.e., effects occurring on electron gyroradius and electron inertial length scales and inverse electron gyrofrequency time scales). The relevant scales are then the ion gyroradius and ion inertial length and inverse ion gyrofrequency scales. Higher frequency ion physics (i.e., ion plasma frequency) that occurs on ion plasma Debye lengths is also excluded, although there are electrostatic hybrid codes that do model such processes. In space plasmas occurring within the Earth's magnetosphere and in the upstream solar wind, the relevant ion-scale lengths are on the order of 10 to \sim 100 km and time scales on the order of seconds. Such time and length scales have been readily resolved by spacecraft over the past 40 years, starting with the ISEE-1,2

spacecraft that first resolved the structure of the Earth's bow shock [3]. Subsequent missions led to the wide use of hybrid codes in interpreting observations and comparing with theoretical models that will become evident in Sect. 3.4. Modern space instrumentation, such as found on MMS, is now able to resolve plasma distributions on much faster time scales associated with some electron phenomena and well resolve the corresponding ion scales. In fact, the four MMS spacecraft have reduced their separation to distances as small as 10 km, where electron dynamics start to become important, and have been able to make measurements down in the magnetic reconnection region [4].

In this section we present a simplified hybrid model with massless electrons, which was presented in the earlier tutorial (Winske et al. [1]), which is still commonly used, that provides a framework for discussion of more complex adaptations later in the chapter. The basic physics model and the associated equations that are solved as a function of space and time are defined as follows. To be consistent with the hybrid model, the ions are treated kinetically using standard particle-in-cell methods [5]. Each simulation ion (with position \mathbf{x}_p , velocity \mathbf{v}_p , charge q_i and mass m_i), which represents many actual physical particles, is subject to the usual (non-relativistic) equations of motion in the electric \mathbf{E} and magnetic \mathbf{B} fields:

$$m_i d\mathbf{v}_p/dt = q_i (\mathbf{E} + \mathbf{v}_p \times \mathbf{B}), \quad (3.1)$$

$$d\mathbf{x}_p/dt = \mathbf{v}_p, \quad (3.2)$$

where the fields have values given on a spatial grid and are interpolated to the location of the simulation particle. The updated particle data for the ion charge density ($q_i n_i$) and current $\mathbf{J}_i = q_i n_i \mathbf{V}_i$ are collected on the grid, from which the number density (n_i) and flow velocity (\mathbf{V}_i) are determined for use later in calculating the electric field. (For simplicity, we assume a single ion species; for multiple species, one accumulates the data for each species separately and then adds them together to get the total ion charge density and current.)

In the standard hybrid model the electrons are treated as a massless fluid, although “inertia-less” is a more accurate description. Thus, the left side of the electron momentum equation is zero, resulting in:

$$d(n_e m_e \mathbf{V}_e)/dt = 0 = -e n_e (\mathbf{E} + \mathbf{V}_e \times \mathbf{B}) - \nabla \bullet \mathbf{P}_e + e n_e \eta \bullet \mathbf{J}, \quad (3.3)$$

where \mathbf{V}_e is the electron fluid velocity, \mathbf{P}_e is the electron pressure tensor, \mathbf{J} is the total current, and η is a resistivity. Even though we assume $m_e = 0$ in Eq. (3.3), we still include the electron pressure term on the right-hand side. This contributes to the electric field that insures that the plasma remains quasi-neutral, and thus eliminating Debye length effects, i.e.,

$$e n_e = q_i n_i, \quad (3.4)$$

where e is the elementary charge. In Eq. (3.3), \mathbf{P}_e is usually taken to be a scalar $\mathbf{P}_e = p_e \mathbf{1}$ but sometimes separate electron pressures along and transverse to the background magnetic field are included, as discussed later in Sect. 3.3. The last term in Eq. (3.3) is an optional resistive term. Usually η is assumed to be a scalar with a constant value. If this term is included, there is a corresponding term $(-e\eta \mathbf{J})$ in the ion equation of motion (3.1) to balance momentum exchange. This term is usually assumed small and is often used as a numerical means to reduce high-frequency fluctuations in the electromagnetic field, but in certain problems (e.g., magnetic reconnection), it can be physically significant, as a localized model for high-frequency waves that initiate reconnection of field lines. We will discuss the pressure model and resistivity later in regard to numerical tests and specific applications.

The electric and magnetic fields in an electromagnetic hybrid model are treated in the low-frequency approximation (i.e., Darwin model [6]) that ignores light waves. Thus, we use Ampere's law,

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} = \mu_0 q_i n_i (\mathbf{V}_i - \mathbf{V}_e), \quad (3.5)$$

and Faraday's law

$$\frac{\partial \mathbf{B}}{\partial t} = -(\nabla \times \mathbf{E}). \quad (3.6)$$

Equation (3.6) is used to advance the magnetic field in time. Equation (3.5) is used to eliminate the electron flow velocity in Eq. (3.3), which then reduces to the equation for the electric field \mathbf{E} . In this low-frequency approximation, the electric field is thus not advanced in time, but instead is determined from the (updated) magnetic field, and ion density and flow velocity. Calculating the electric field at the next time step is the essential problem in all hybrid algorithms and different methods to accomplish this will be discussed throughout this chapter.

The numerical implementation of these equations in a hybrid code begins by choosing a spatial grid. Often, but not always, a rectangular grid is used. One must also decide on the location of fields and particle source terms on the grid. In the basic implementation we describe here, the electric field, ion density, ion current, and electron pressure are located on the vertices of the grid, while the magnetic field is located at the center of the grid [7]. In this way, curl \mathbf{E} is given correctly at the location of \mathbf{B} and curl \mathbf{B} at the location of \mathbf{E} . Given the grid, the ion dynamics are done as in usual PIC codes: the fields are interpolated to the positions of the particle ions to give the correct local force and thus acceleration of the ions, and after the particles are moved, their contributions update the ion density and current values at the grid points. This can be done using linear, or another type of, weighting. While different hybrid algorithms may choose different ways to locate the fields and collected ion source terms on the grid, we will not discuss these methods further in this chapter. The ways that this can be done are similar to those used in PIC simulations that are described in Chap. 6. However, it should be noted that some

recent work has suggested that spatial resolution, particle shapes, and smoothing in hybrid codes can affect low-frequency wave dispersion [8]. Thus, one always needs to apply suitable test problems to any hybrid code one is using.

The advance in time also follows PIC methodology and proceeds as follows. We start at time step N (denoted by superscript), when the particle positions \mathbf{x}_p^N and the electric field \mathbf{E}^N and magnetic field \mathbf{B}^N are known, while the particle velocities are a half-time step behind, $\mathbf{v}_p^{N-1/2}$. This allows the particle positions and velocities to “leapfrog” in time. First, the particle velocities are advanced to the next level $N+1/2$ with time step Δt in the equations of motion:

$$\mathbf{v}_p^{N+1/2} = \mathbf{v}_p^{N-1/2} + \frac{q_i}{m_i} (\mathbf{E}^N + \mathbf{v}_p^N \times \mathbf{B}^N) \Delta t, \quad (3.7)$$

And then the particle positions are similarly advanced to time level $N+1$

$$\mathbf{x}_p^{N+1} = \mathbf{x}_p^N + \mathbf{v}_p^{N+1/2} \Delta t. \quad (3.8)$$

In the process the ion current is collected at time step $N+1/2$ and the density at $N+1$. We also save the old value of the ion density n_i^N for later use. We note that in Eq. (3.7), $\mathbf{v}_p^N = \frac{1}{2}(\mathbf{v}_p^{N-1/2} + \mathbf{v}_p^{N+1/2})$, which can then be solved to get $\mathbf{v}_p^{N+1/2}$, e.g. [9].

Next, the magnetic field \mathbf{B}^N is advanced to $\mathbf{B}^{N+1/2}$ using \mathbf{E}^N and Faraday’s law, Eq. (3.6):

$$\mathbf{B}^{N+1/2} = \mathbf{B}^N - \frac{\Delta t}{2} (\nabla \times \mathbf{E}^N). \quad (3.9)$$

By solving the electron momentum Eq. (3.3) for the electric field, we can evaluate $\mathbf{E}^{N+1/2}$:

$$\begin{aligned} \mathbf{E}^{N+1/2} &= -\mathbf{V}_i^{N+1/2} \times \mathbf{B}^{N+1/2} - \frac{\nabla p_e^{N+1/2}}{q_i n_i^{N+1/2}} - \frac{\mathbf{B}^{N+1/2} \times (\nabla \times \mathbf{B}^{N+1/2})}{\mu_0 q_i n_i^{N+1/2}} \\ &= \mathbf{E}^{N+1/2} \left(\mathbf{B}^{N+1/2}, \mathbf{V}_i^{N+1/2}, n_i^{N+1/2} \right), \end{aligned} \quad (3.10)$$

because we know the ion current $\mathbf{V}_i^{N+1/2}$ from collecting the new ion particle velocities $\mathbf{v}_p^{N+1/2}$, the ion density from averaging its old and new values, $n_i^{N+1/2} = \frac{1}{2}(n_i^N + n_i^{N+1})$, and $\mathbf{B}^{N+1/2}$ from Eq. (3.9). On the far right side of Eq. (3.10) we have explicitly written out the time levels of the appropriate variables needed to compute \mathbf{E} , for later reference. We have also assumed an adiabatic form for the scalar electron pressure:

$$p_e = n_o T_{eo} (n_e/n_o)^\gamma, \quad (3.11)$$

where T_{eo} (n_o) is a reference electron temperature (density), usually with $\gamma = 5/3$ (adiabatic), so that from Eq. (3.4), p_e^N involves just knowing n_i^N .

Then again, we use Faraday's law (3.6), as in Eq. (3.9) with $\mathbf{B}^{N+1/2}$ and $\mathbf{E}^{N+1/2}$ to advance the magnetic field to \mathbf{B}^{N+1} . However, the advance of $\mathbf{E}^{N+1/2}$ to \mathbf{E}^{N+1} is more problematic, as noted earlier. An examination of applying Eq. (3.10) for \mathbf{E}^{N+1} shows that while \mathbf{B}^{N+1} and n_i^{N+1} are known, \mathbf{V}_i^{N+1} is not.

Our earlier tutorial [1] described a number of ways in which this issue could be resolved. Two of those methods, namely the predictor-corrector approach [10] and the Current Advance Method and Cyclic Leapfrog (CAM-CL) approach [11], are discussed in some detail in that book chapter. Both of these methods are actively used in present-day hybrid simulation codes and will be discussed later in Sect. 3.4. These two methods, along with a third method discussed here, will be compared later in Sect. 3.5.3. Moreover, a recent extension of the predictor-corrector method is presented in Sect. 3.5. In this updated tutorial, we restrict ourselves here to a third, and the simplest, method to advance the electric field its final half-time step. In this approach, before we advance the particles to calculate $\mathbf{v}_p^{N+1/2}$ and thus accumulate particle information on the grid to determine $\mathbf{V}_i^{N+1/2}$, we retain $\mathbf{V}_i^{N-1/2}$ (as well as $q_i n_i^N$) and then use a simple linear extrapolation to obtain:

$$\mathbf{V}_i^{N+1} = \frac{3}{2}\mathbf{V}_i^{N+1/2} - \frac{1}{2}\mathbf{V}_i^{N-1/2}. \quad (3.12)$$

Then \mathbf{E}^{N+1} can be evaluated using Eq. (3.10) and the update to time level $N+1$ is completed. A flow chart of the overall time advance process for this velocity-extrapolation method is given in Fig. A1 of Karimabadi et al. [12]. Note that in this method, as well as in others discussed later [10, 11], the determination of the electric field does not require a global elliptic solve, so that hybrid codes that use these methods are readily scalable to run on very large computer architectures.

In the hybrid model the shortest wavelength modes in the system are whistlers, which because of the quadratic dispersion of the wave frequency ω in terms of the wavenumber k , $\omega \sim k^2$ for parallel propagation can be problematic. The issue can be minimized (without adding resistivity) to some degree by using a smaller time step to just advance the magnetic field from N to $N+1/2$, while retaining a larger time step for the ions. This sub-stepping is done using Faraday's law, Eq. (3.6), where \mathbf{E} is expressed in terms of its source terms, Eq. (3.10):

$$\mathbf{B}^{N+1/2} = \mathbf{B}^N - \frac{\Delta t}{2} \nabla \times \mathbf{E}^N (\mathbf{B}^N, \mathbf{V}_i^N, n_i^N). \quad (3.13)$$

In this case a fourth order Runge-Kutta scheme [13] is used, where on the right-hand side of Eq. (3.13) we keep n_i^N and \mathbf{V}_i^N fixed and just advance \mathbf{B} . Specifically, we subcycle \mathbf{B} from N to $N+1/2$ in L sub-steps so that $\Delta t' = (\Delta t/2)/L$. At the l -th sublevel, $l = 1, 2, \dots, L$, with $N' = N + l/L$, we advance \mathbf{B} to $N'+1/L$ as follows:

$$\mathbf{B}^{N'+1/L} = \mathbf{B}^{N'} + \frac{\Delta t'}{6} \sum_{j=1}^4 f_j \mathbf{K}_j^{N'}, \quad (3.14)$$

where

$$\mathbf{K}_j^{N'} = -\nabla \times \mathbf{E} \left(\mathbf{B}^{N'} + g_j \Delta t' \mathbf{K}_{j-1}^{N'}, \mathbf{V}_i^N, n_i^N \right), \quad (3.15)$$

and $f_j = (1, 2, 2, 1)$, $g_j = (0, \frac{1}{2}, \frac{1}{2}, 1)$ and $\mathbf{K}_0 = 0$.

This completes the discussion of a basic hybrid code algorithm that neglects electron inertia. It is the basis of the codes used in the applications discussed in the next section Sect. 3.3. This is followed by Sect. 3.4 with ongoing work based on well-known hybrid codes that use different hybrid algorithms and by Sect. 3.5 on the development of new methods that expand the range of problems that can be studied.

3.3 Examples of Current Hybrid Code Applications

Next we discuss recent applications of large-scale hybrid codes to problems of current interest in space physics. In each case the phenomena of interest can occur over significant regions of space—typically greater than several Earth radii (R_E). On these scales, full electron and ion dynamics cannot be modeled, except perhaps by using very small ion to electron mass ratios. On the other hand, the physical processes involve significant ion kinetic phenomena so that an MHD or Hall-MHD model is also not sufficient. Large-scale hybrid codes, either in 2D or 3D, can now capture the essential physics on scales that can be directly related to spacecraft measurements. We consider three examples: one involving phenomena upstream of planetary bow shocks, one concerning magnetic reconnection in the magnetotail, and a third illustrating more global processes centered on the Earth’s magnetosheath, but extending out toward the bow shock and in toward the magnetopause as well. All of the hybrid codes in this section use some form of velocity extrapolation in computing the advanced-time electric field.

3.3.1 Planetary Foreshocks and Bow Shocks

Global hybrid simulations and multi-spacecraft observations have resulted in a dramatic increase in our knowledge of planetary ion foreshock processes and their impacts on the magnetosheath and magnetosphere/ionosphere system responsible for the formation of the bow shock [14]. These processes involve the generation of low-frequency waves driven by back-streaming ions in the foreshock as well as discontinuities in the solar wind and their interaction with the shock. The scale lengths of these structures extend from several ion inertial lengths (d_i) to $\sim 1 R_E$ ($R_E/d_i \sim 60$) at Earth—scales that are easily resolved by hybrid simulations. A simulation of such processes involves the magnetized solar wind plasma flowing in one end of the simulation box toward a stationary plasma and magnetic field that represents the planet, with appropriate inflow boundary conditions on the particles and fields. Typically, the simulation runs for some period of time to set up the bow

shock and foreshock self-consistently and allows the solar wind to flow through and around the shock, with corresponding outflow boundary conditions on the other sides of the simulation. We present an illustration of this technique recently applied to the solar wind interaction with Venus.

Global hybrid simulations with Alfvén Mach number ($M_A = \text{solar wind speed}/\text{Alfvén speed}$) greater than three show the presence of structures in the Earth's foreshock consisting of cores of reduced density and magnetic field and rims of enhanced field and density, termed foreshock cavitons [14, 15]. The size of these cavitons corresponds to $\sim 1 R_E$ at Earth, implying that many foreshock cavitons can fit into the Earth's foreshock [16]. The results of the 3D global hybrid simulations presented here show the formation of foreshock cavitons also occurs at Venus despite the much smaller size of its foreshock [17]. In these calculations the box size is $400 \times 300 \times 300 \text{ cm}^3/\omega_{pi}$ with cell size $\Delta x = \Delta y = \Delta z = 1 \text{ cm}/\omega_{pi}$ (where $c/\omega_{pi} = \text{ion inertial length} = d_i$), which for typical solar wind values is $\sim 100 \text{ km}$. The model uses a total of 6×10^8 particles to represent the solar wind protons and ionospheric O⁺ ions and a time step of $0.0025 \Omega_p^{-1}$ where Ω_p is the proton gyrofrequency. Figure 3.1 shows results from such a run, with panel (a) corresponding to ion temperature, log of density in panel (b), and the total magnetic field strength in the bottom panel. In these plots X is the direction of the solar wind flow, the magnetic field lies in the X-Y plane. Examples of foreshock cavitons are evident in the figure; their size is comparable to the radius of Venus. The figure also shows the presence of a foreshock compressional boundary (FCB) associated with enhancements in density and magnetic field at the edges of the foreshock [18, 19]. As with the underlying waves forming them, foreshock cavitons are carried by the solar wind toward the bow shock. As they approach the shock, their density and magnetic field rims become larger and the core region becomes associated with highly decelerated and heated solar wind [20].

3.3.2 Improved Electron Pressure Closure for Magnetic Reconnection in the Magnetosphere

As discussed earlier, a simple isotropic, adiabatic electron pressure model is often assumed in hybrid codes. However, for applications such as for magnetic reconnection that commonly occurs at the magnetopause and in the magnetotail (and even in the magnetosheath and upstream of the bow shock, as we will see shortly), a more detailed model may be necessary to improve agreement with theory and observations. In this sub-section, we review recent results from implementing an improved electron fluid closure in hybrid simulations. The application considered—a single reconnecting current sheet—is relatively simple. Observational, theoretical, and fully kinetic numerical studies of magnetic reconnection have demonstrated that strong electron pressure anisotropy, with different effective temperatures in the directions parallel and perpendicular to the magnetic field, may develop in current sheets in space [21]. Most importantly for hybrid modeling, this electron pressure anisotropy is not confined to small electron-scale boundary layers. Rather,

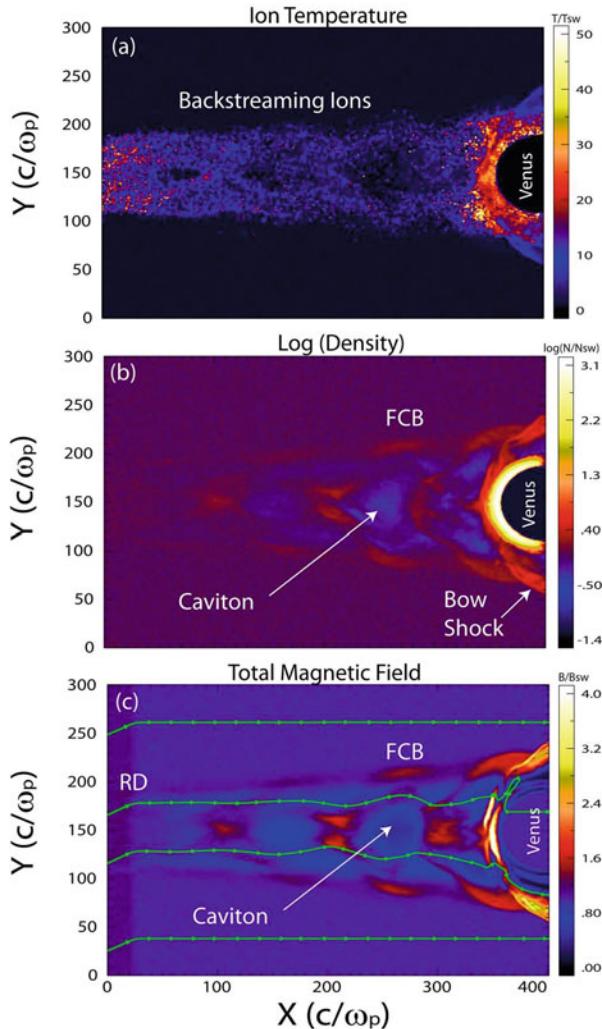


Fig. 3.1 Plots of ion temperature, density and magnetic field strength showing the structure of the Venusian foreshock. Also shown in panel (c) are magnetic field lines indicating the presence of a rotational discontinuity (RD) in the solar wind

the electron pressure tensor may remain anisotropic over scales greater than tens of ion inertial lengths and couple back to the ion-scale system evolution.

A main mechanism for generating electron pressure anisotropy near reconnection layers is the trapping of electrons in an effective ambipolar quasi-potential $\Phi_{||}$, which is the parallel electric field integrated along the magnetic field lines. This trapping process is well-understood in the guiding center picture, and a relatively simply model has been formulated to describe the electron velocity distribution

when there is a guide magnetic field present (so that the electron particle orbits remain magnetized throughout the reconnecting current sheet) [22]. From the kinetic model, a set of anisotropic equations of state (AEoS) for the electron pressure tensor components parallel and perpendicular to the magnetic field may be derived [23]. The AEoS are similar to the Chew-Goldberger-Low (CGL) fluid closure, which gives the pressure components in terms of the density and magnetic field strength. Indeed, the AEoS reduce to the CGL scaling in the limit of strong electron trapping.

The AEoS have been implemented in reconnection models in the hybrid code H3D [24] by including the electron pressure tensor divergence in the Ohm's law for the electric field. The AEoS hybrid model has recently been validated with observations gathered by NASA's MMS mission [2] at a magnetopause reconnection site [25]. In Fig. 3.2, the electron pressure tensor components observed by MMS plotted here as electron density and temperatures parallel and perpendicular to the local magnetic field show remarkably good agreement with the AEoS predictions from a

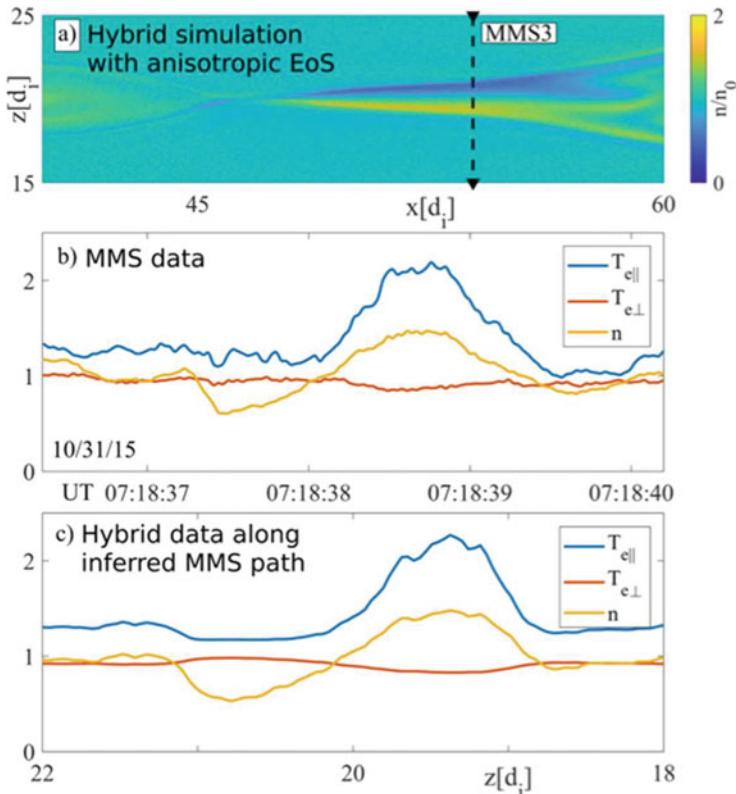


Fig. 3.2 (a) Plasma density from a hybrid simulation matching the parameters of an MMS observation of a reconnection exhaust. (b) MMS measurements of the plasma density and electron temperature components. (c) Data from the hybrid model along the cut shown in (a) displaying good agreement with the MMS observations

hybrid model setup to match the plasma parameters of this reconnection event. Note that the hybrid model with AEoS, which assume a very fast parallel transit time of electrons, reproduces this electron anisotropy even more accurately than a fully kinetic simulation performed with a reduced mass ratio of $m_i/m_e = 100$.

The MMS event considered above contained a fairly strong guide magnetic field $B_g/B_0 > 1$. The effect of the electron pressure anisotropy is more pronounced in other regimes [26, 27] with weaker guide magnetic fields, when the electron beta is high near the X-line and the electron pressure becomes more dynamically important. For upstream electron betas of a few percent and guide magnetic fields of $B_g/B_0 \sim 0.2\text{--}0.6$, a regime exists where the electron pressure is close to the firehose threshold, $p_{||} - p_{\perp} \sim B^2/\mu_0$, throughout the exhaust region, meaning the electron pressure anisotropy nearly balances the magnetic tension forces on the electron fluid. This allows strong perpendicular electron currents to flow, forming elongated current sheets extending from the X-line. These extended electron current sheets, as illustrated in Fig. 3.3, reach across ion scales and are limited in the simulations only by the size of the numerical domain.

3.3.3 Magnetosheath: Effects of the Bow Shock, Turbulence, and Reconnection

Due to the availability of in situ spacecraft measurements, the Earth's bow shock provides an important example for basic studies of fast magnetosonic shocks and turbulence. While observations provide a local snapshot of the plasma, simulations provide a larger scale as well as dynamical view of the environment. This complementary approach has proven quite effective in uncovering fundamental physics of collisionless shocks and associated turbulence. In this sub-section we discuss global hybrid simulations of the solar wind interacting with the Earth's magnetosphere, as described in detail in Karimabadi et al. [28]. Details of the hybrid code used in these calculations are found in Karimabadi et al. [29]. While results from the simulations displayed as individual figures give the reader some sense of the complex interactions that are occurring, computer-generated movies of the simulations discussed in this sub-section are also available with links given in the text.

In these 2D simulations, as with those reported earlier in Sect. 3.1, the bow shock, as well as the foreshock region upstream and the magnetosheath region downstream, are formed self-consistently in the simulation. The solar wind is uniformly and continuously injected from the left boundary and interacts on the right with the planetary magnetic field, a line dipole. The simulation is run for a long period of time in order to remove initial transients. Then after some period of time, a rotational discontinuity is introduced at the left boundary by changing the direction of the interplanetary magnetic field.

While Karimabadi et al. [28] considered the properties of magnetosphere under a variety of conditions, the focus here is on their resolution to a key issue that was not previously reconciled between observations and simulations. This issue

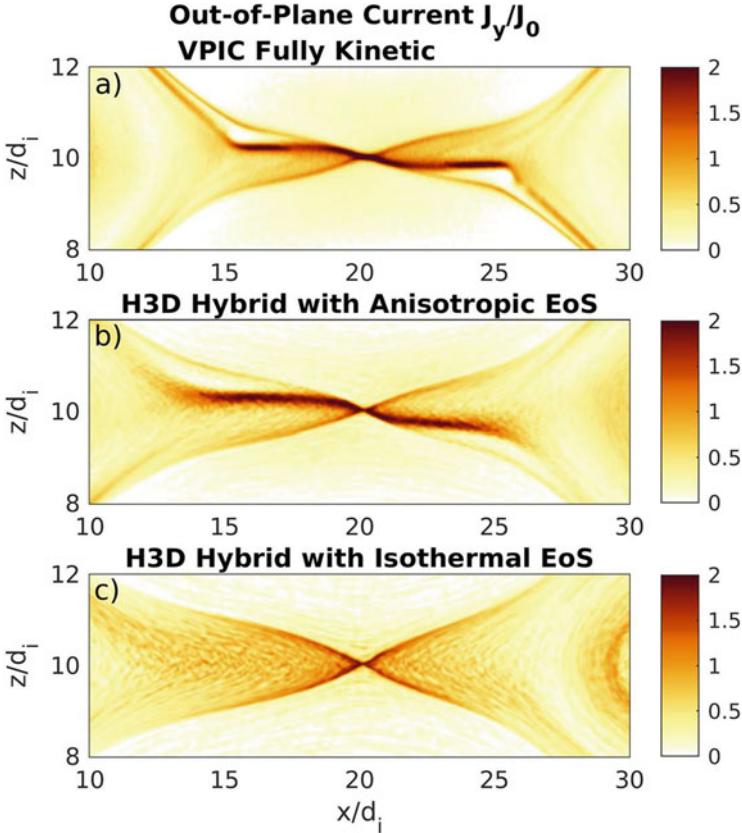


Fig. 3.3 Out-of-plane current density in a fully-developed reconnection region from (a) a fully kinetic simulation, (b) a hybrid simulation including electron pressure anisotropy, and (c) a hybrid simulation with isothermal electron closure. The electron pressure anisotropy supports the elongated current layer in panels (a) and (b)

was related to observations of a small reconnection event in the magnetosheath by Retino et al. [30], which they attributed to the effect of turbulence. It remained unclear whether this was just an isolated, anomalous event or a common feature of the magnetosheath turbulence. And it remained a mystery as to why no kinetic simulation of the quasi-parallel shocks, either in isolation or in a global kinetic simulation of the magnetosphere, had seen evidence for such reconnection events. Karimabadi et al. [28], using a series of peta-scale hybrid simulations and advanced visualization techniques, were able to resolve these questions. In the first simulation shown here, the system size is 8192×8192 ion inertial lengths (d_i) with cell size of $1.0 d_i$ and 200 particles per cell. Figure 3.4a shows an intensity plot of the ion density of a segment of the simulation that clearly demonstrates the formation of waves in the foreshock and a portion of the magnetosheath. In the magnetosheath,

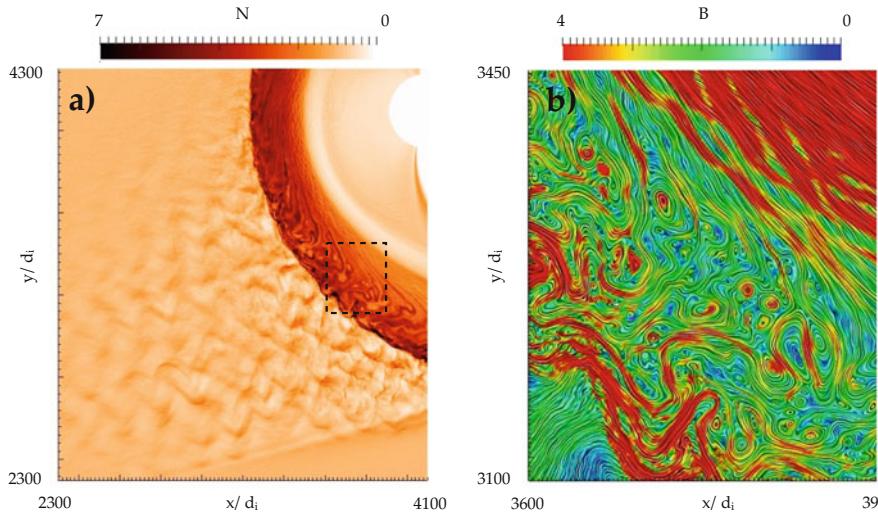


Fig. 3.4 Foreshock and magnetosheath turbulence in a 2D hybrid simulation describing interaction of solar wind injected from the left boundary with a dipolar field. (a) ion density; (b) LIC visualization of the magnetic field structure in a sub-region marked on the left panel. The scales are normalized to the ion inertial length d_i

current sheets and magnetic islands can also be seen. The magnetic islands are a consequence of reconnection events, first reported by Retino et al. [30], which serve as an important dissipation mechanism.

Figure 3.4b shows a close-up of the quasi-parallel magnetosheath. The Line Integral Convolution (LIC) technique, which is a unique method to represent streamlines of a vector field that is described in the appendix of Karimabadi et al. [28], is used to illustrate the structure of the magnetic field B with the color corresponding to the magnitude of B . Many magnetic islands are observed near the shock surface and filling the magnetosheath to the vicinity of the magnetopause. The dynamical evolution of the magnetosphere in this case as well as a zoomed in area of the simulation around the quasi-parallel magnetosheath is shown in two movies at <https://doi.org/10.7910/DVN/L7D4CP> [S1_den.mp4] and [b1_lic.mp4], respectively. These findings demonstrated that small-scale reconnection events are a common by-product of the turbulence developed downstream of quasi-parallel shocks, thus establishing a link between shocks, turbulence, and reconnection. Subsequent high-resolution spacecraft observations with MMS have confirmed the presence of reconnection events in the magnetosheath and have strengthened their association with turbulence [31, 32].

3.4 Other Hybrid Algorithms, Codes, and Applications

In the previous section, we showed examples of the use of 2D and 3D hybrid codes applied to recent problems in magnetospheric physics. All of these calculations use algorithms based on the extrapolation of the ion velocity in order to advance the electric field to the next time step. In this section, we first give brief examples of other hybrid codes based on different algorithms that use alternative methods to advance the electric field and how these codes are being applied. We include a brief description of hybrid modes that include finite-electron inertia. We then discuss in a bit more detail several new algorithms that are designed to deal more carefully with issues related to energy and momentum conservation. We conclude this section with comments related to comparison of different hybrid algorithms.

The solar wind is a very dynamic plasma environment, especially in the kinetic ion regime for which hybrid codes provide optimal capabilities. Hybrid codes used by Liewer et al. [33], Hellinger et al. [34–37], and Franci et al. [38–40] all employ the CAM-CL algorithm developed by Matthews [11]. It contains two significant changes in the calculation of the electromagnetic fields relative to the hybrid algorithm presented in Sect. 3.2. First, the advance from time step N to time step $N+1$ again involves a slight approximation. In the preceding discussion related to Eq. (3.12), an extrapolation to advance $\mathbf{V}_i^{N+1/2}$ to \mathbf{V}_i^{N+1} was used in order to advance the electric field to the next time step, \mathbf{E}^{N+1} . Instead, in the CAM-CL approach the equation for the ion velocity moment is used:

$$\frac{\partial \mathbf{V}_i}{\partial t} = \frac{q_i}{m_i} (\mathbf{E} + \mathbf{V}_i \times \mathbf{B}) - \frac{1}{n_i m_i} \nabla \bullet \mathbf{P}_i. \quad (3.16)$$

The right-hand side of this equation contains the next higher velocity moment, namely the ion pressure tensor \mathbf{P}_i , which is not needed (or calculated) in a hybrid algorithm. Instead, \mathbf{V}_i^{N+1} is approximated by ignoring this term, i.e., by keeping only the “free-streaming” contribution: $(q_i/m_i)(\mathbf{E}^{N+1/2} + \mathbf{V}_i^{N+1/2} \times \mathbf{B}^{N+1/2})$. The second change in this algorithm is to keep two copies of the magnetic field, one associated with integer time steps ($N, N+1, \dots$), the other with half-time steps ($N+1/2, N+3/2, \dots$) that are leap-frogged over each other. As before, each version of \mathbf{B} is advanced using Faraday’s law with smaller sub-steps. These two differences, i.e., using a moment approach to update the current and the leap-frogged magnetic field advance, explains why this method is referred to as “Current Advance Method and Cyclic Leapfrog” (CAM-CL). Matthews [11] provides a flow diagram (his Fig. 1) to better explain his algorithm.

To study the development of low-frequency instabilities in the solar wind, a moving box technique developed by Liewer et al. [33] was used by Hellinger et al. [34] to follow a parcel of solar wind as it flows out from the sun. As a consequence of the expansion of the solar wind, the background magnetic field and density fall off with radial distance. In the simulations this effect is included as an extra force on the ions that leads to a cooling of the expanding plasma and the development of ion temperature anisotropies. Using this technique, Hellinger and Trávníček [35]

have studied the growth and nonlinear development of parallel and oblique firehose instabilities showing how these two modes effectively limit the ion temperature anisotropy. In the presence of large-scale Alfvénic turbulence characteristic of the solar wind, the development and long-time behavior of both firehose ($T_{i\parallel} > T_{i\perp}$) and mirror ($T_{i\perp} > T_{i\parallel}$) instabilities, and their overall effects on the wave turbulence and the ion temperature, combine in a rather complex manner [36, 37].

The CAM-CL algorithm has also been used by Franci et al. [38–40] in their 2D/3D CAMELIA hybrid code to study the development and evolution of plasma turbulence in the solar wind. Here, as opposed to outward expanding plasma in Hellinger et al. [36], the plasma is at rest and effects associated with the expansion of the solar wind are neglected. In this case the turbulence is applied as long wavelength magnetic field fluctuations at $t = 0$, and then the system evolves to shorter wavelengths in time, eventually reaching a quasi-steady state. The power spectrum of the perpendicular magnetic fluctuations, $P(k_\perp)$, in this stage varies at long wavelengths as $k_\perp^{-5/3}$, consistent with fluid turbulence. In regions corresponding to ion scales (i.e., ion inertial lengths or ion gyroradii), the power spectrum changes and varies as k_\perp^{-3} , characteristic of kinetic Alfvénic wave turbulence. The 2D simulations [39] are also used to examine how the number of simulation particles per cell and the size of the applied (constant) resistivity change the results and indicate where in parameter space the results are determined by physical properties and not by numerical quantities. With these effects having been quantified, 3D simulations [40] can then be done and compared with the 2D results, using similar initial conditions. It is found that similar spectral indices occur in perpendicular magnetic fluctuations, but smaller indices are found for density and velocity fluctuations in 3D, suggesting the importance of compressional effects.

Other hybrid codes are in common use for a wide variety of applications. For example, Brecht et al. [41] have carried out 2D and 3D simulations of the solar wind interaction with Mars and also Venus. Their code HALFSHEL uses the predictor-corrector algorithm of Harned [10], which will be discussed later in this section. The solar wind interaction with Mars, for example, is complicated by the eccentricity of the Mars orbit, the variability of the solar wind, and the presence of crustal magnetic fields, which make for challenging interpretations of observations from the MAVEN spacecraft [42]. Very good comparisons between the 3D hybrid simulations and MAVEN data concerning the ion loss rate from the planet have been carried out by Ledvina et al. [43]. Another example of a solar wind interaction is with the surface of the moon, as studied by Lipatov et al. [44] using 3D hybrid simulations. In this case various ion species (e.g., He^+ , He^{++} , Na^+) are excited, escape from the moon's surface, and are picked up by the solar wind. The boundary conditions that model the conducting surface of the moon are also included in the model. The properties of the various ion species that are picked up by the solar wind have been compared with fly-by observations of the moon from the ARTEMIS spacecraft [45]. Continuing in the planetary vein, the work by Trávníček et al. [46] regarding the interaction of the solar wind with Mercury should also be acknowledged, especially in regard to the excellent interpretation of the data from flybys and then orbits of the planet by the Mercury MESSENGER spacecraft [47]. For example, their hybrid simulations

(using the CAM-CL method) are able to explain the differences in waves seen upstream and downstream of the bow shock and the absorption or escape of solar wind ions from the planet. Other simulations have investigated the differences in the interaction with Mercury's magnetosphere with the solar wind for both northward and southward interplanetary magnetic field [48].

The CAM-CL algorithm has also been used in 2D and 3D hybrid simulations to investigate in detail the structure of high Alfvén Mach number ($M_A > 3$), perpendicular (upstream magnetic field perpendicular to the shock normal), collisionless shocks by Burgess et al. [49]. Their very finely resolved simulations have studied in detail the ion-driven waves in the ramp region just upstream of the shock ramp. In particular, the inclusion of the third spatial dimension allows fluctuations propagating parallel to the magnetic field to interact with ions reflected by the shock to create a complex pattern of obliquely propagating waves. Shock waves caused by coronal mass ejecta overtaking a slow portion of the solar wind have been studied by Gargaté et al. [50], using their 2D/3D hybrid code dHybrid [51], and subsequently employed by Caprioli and Spitkovsky [52] to model astrophysical shock waves. The structure of shock waves has also recently been studied with a new hybrid code similar to dHybrid, developed by Vshivkova et al. [53]. Over the years Giacalone and colleagues have used hybrid codes (based on velocity extrapolation) to investigate particle acceleration at shocks. An example of their recent work is a 3D simulation study of thermal proton acceleration at parallel shocks [54].

The interaction of the solar wind with the Earth's magnetosphere has been studied for a number of years by Lin et al. [55], based on an algorithm originally developed by Swift [56]. This code is unique in that it uses a non-rectangular grid that conforms better to the shape of the bow shock and the magnetopause. The field solver uses a combination of extrapolation and predictor-corrector techniques. This code has been recently used to understand the generation of short-wavelength kinetic Alfvén waves in the high latitude magnetotail due to the excitation of shear Alfvén waves by fast Earthward flows during substorms [57].

Finally, while we have emphasized hybrid models in which electron inertia is neglected, with modern computer techniques also hybrid-kinetic models that retain electron inertia can efficiently be run. Much interest and development has occurred in these areas, especially in recent years, motivated by new space observations (e.g., MMS) and laboratory experiments of magnetic reconnection. An early type of finite-mass hybrid code for laboratory plasmas was developed decades ago [58], and more recently an algorithm based on a reduced set of electromagnetic field equations [59] was used to simulate small-scale structures associated with the electron diffusion region in magnetic reconnection. This type of reconnection physics is usually investigated with large-scale fully kinetic particle codes [60]. Discussion of issues unique to these types of hybrid algorithms is found in the monograph by Lipatov [61] and in Winske et al. [1]. Recently, interest in finite-electron mass hybrid algorithms has resurfaced due to the development of new fluid electron models included in a generalized quasi-neutral model by Amano [62] and in a more rigorous method for including electron inertia by Muñoz et al. [63], which is the basis of the CHIEF code. These models are described in detail in Chaps. 9 and

10 of this book. There is also continued interest in hybrid and fully kinetic Vlasov codes (Chap. 5 of this book) as well as in fully kinetic PIC codes explicit (Chap. 6 of this book) as well as implicit one (see, e.g., [64, 65] and Chap. 8 of this book).

3.5 New Hybrid Algorithms

While most of the discussion in this section thus far has addressed relatively recent results obtained with hybrid algorithms that are based on methods that have been around for decades, there have been recent and important developments in new algorithms. These new efforts are driven by the need to improve the accuracy of hybrid calculations that are now able to run on many thousands of processors with billions (and even trillions) of particles for on the order of a million time steps. In such very large, time-consuming, and expensive calculations, the issues of energy conservation and slow-growing numerical instabilities become more important. It requires a more careful examination of how the particle information is transferred to the grid and how the electromagnetic fields are solved, with the ever-present constraint of hybrid models that the electric field is not advanced in time. Here we discuss two very different approaches to address these questions. In addition, the discussion of a third, relatively new code, HYPERS [66], developed by Omelchenko and Karimabadi that includes asynchronous time advancement, is considered in more detail in Chap. 13.

3.5.1 An Advanced Hybrid Model

One of the new hybrid codes, PEGASUS, has been developed by Kunz et al. [67] specifically for astrophysical applications. The code is unique in that additional force terms (i.e., gravity, Coriolis and centrifugal) have been introduced to model rotating and shearing plasmas that occur in astrophysics, using a shearing sheet formulation.

The code also uses a refined predictor–predictor–corrector algorithm to more accurately advance the electric and magnetic fields to the next time step. To explain this procedure, it is useful to recall the traditional predictor–corrector scheme used in hybrid codes [10] in order to obtain the advanced electric field \mathbf{E}^{N+1} . As shown earlier, Eqs. (3.1–3.10), the ion velocities and the magnetic field can be advanced to time level $N+1/2$, and the ion density can also be determined at that time, which means that $\mathbf{E}^{N+1/2}$ can also then be calculated. A predicted value for the electric field \mathbf{E}'^{N+1} can be readily obtained from the time-centered equation of the form,

$$\mathbf{E}^{N+1/2} = \frac{1}{2} (\mathbf{E}^N + \mathbf{E}'^{N+1}). \quad (3.17)$$

by rearranging:

$$\mathbf{E}'^{N+1} = 2\mathbf{E}^{N+1/2} - \mathbf{E}^N. \quad (3.18)$$

Using this \mathbf{E}'^{N+1} in Faraday's law (3.6), a predicted magnetic field \mathbf{B}'^{N+1} is then obtained from the known field $\mathbf{B}^{N+1/2}$. The particles are then advanced ahead in these predicted fields an additional time step in order to collect the moments, $n_i'^{N+3/2}$ and $\mathbf{V}_i'^{N+3/2}$. Then again using Faraday's law with \mathbf{E}'^{N+1} and \mathbf{B}'^{N+1} , $\mathbf{B}'^{N+3/2}$ is obtained. From $\mathbf{B}'^{N+3/2}$, $n_i'^{N+3/2}$ and $\mathbf{V}_i'^{N+3/2}$, $\mathbf{E}'^{N+3/2}$ is computed as in Eq. (3.10), and then the corrected \mathbf{E}'^{N+1} can be obtained in a symmetric fashion:

$$\mathbf{E}^{N+1} = \frac{1}{2} (\mathbf{E}^{N+1/2} + \mathbf{E}'^{N+3/2}). \quad (3.19)$$

With this advanced electric field and $\mathbf{B}^{N+1/2}$, \mathbf{B}^{N+1} is then computed using Faraday's law. The flow chart in Fig. A1 of Karimabadi et al. [12] nicely summarizes the overall algorithm.

In the predictor–predictor–corrector scheme of Kunz et al. [68], instead of using Eq. (3.18), one looks for a more time-centered approach so that the desired predicted quantity appears on the left side of Eq. (3.17), while the quantities on the right-hand side are either known or previously predicted. Thus, they would consider the predicted electric field of the form (Eq. 3.18) as an unacceptable starting point. In their formulation, \mathbf{v}_p^N and \mathbf{x}_p^N are known at the same time, so the equations look a bit different, but the idea is similar. An overview of the process described here is shown schematically in their Fig. 2. In this case, a predicted \mathbf{B}'^{N+1} is obtained from Faraday's law with \mathbf{E}^N . A predicted \mathbf{E}'^{N+1} is computed from \mathbf{B}'^{N+1} , n_i^N and \mathbf{V}_i^N (i.e., computed using Eq. (3.10), but with quantities at mixed time levels) and then $\mathbf{E}'^{N+1/2}$ is obtained in a time-centered manner:

$$\mathbf{E}'^{N+1/2} = \frac{1}{2} (\mathbf{E}^N + \mathbf{E}'^{N+1}), \quad (3.20)$$

and $\mathbf{B}'^{N+1/2}$ is determined from \mathbf{B}^N and \mathbf{B}'^{N+1} in the same manner as (Eq. 3.20).

The particles are then advanced to time level $N+1$ using $\mathbf{E}'^{N+1/2}$ and $\mathbf{B}'^{N+1/2}$ to obtain the estimated moments $n_i'^{N+1}$ and $\mathbf{V}_i'^{N+1}$, which along with \mathbf{B}'^{N+1} allows computing a second predictor electric field \mathbf{E}''^{N+1} , using Eq. (3.10). In addition, from $\mathbf{E}'^{N+1/2}$, $\mathbf{B}'^{N+1/2}$ and Faraday's law, \mathbf{B}''^{N+1} is computed. Then again, in the time-centered form, one obtains a second predicted electric field at time level $N+1/2$:

$$\mathbf{E}''^{N+1/2} = \frac{1}{2} (\mathbf{E}^N + \mathbf{E}''^{N+1}); \quad (3.21)$$

and likewise $\mathbf{B}''^{N+1/2}$ is obtained from \mathbf{B}^N and \mathbf{B}''^{N+1} . $\mathbf{E}''^{N+1/2}$ and $\mathbf{B}''^{N+1/2}$ are then used to push the particles ahead to time level $N+1$, obtaining \mathbf{v}_p^{N+1} and \mathbf{x}_p^{N+1} .

In addition, the moments n_i^{N+1} and \mathbf{V}_i^{N+1} are obtained, and also \mathbf{B}^{N+1} that is computed from $\mathbf{B}''^{N+1/2}$ and Faraday's law with $\mathbf{E}''^{N+1/2}$. Then finally \mathbf{E}^{N+1} can be computed since n_i^{N+1} , \mathbf{V}_i^{N+1} , and \mathbf{B}^{N+1} are all known.

Kunz et al. [67] have carried out a number of numerical tests to validate their new code, including Landau damping, linear Alfvén and whistler waves, parallel shocks, and a magneto-rotational instability. PEGASUS has also been used to study the excitation and nonlinear evolution of the firehose ($p_{i\parallel} > p_{i\perp}$) and mirror ($p_{i\perp} > p_{i\parallel}$) instabilities ($p_{i\parallel,\perp}$ = ion pressure, along or transverse to the magnetic field) in a plasma with a persistent linear shear, such as occurs in the solar wind or galactic accretion flows [68]. In the case where the magnetic field magnitude decreases so that $p_{i\perp}$ decreases, the firehose instability is excited and saturates at a low level, consistent with the stability threshold, due to ion scattering. In the opposite case of an increasing magnetic field and thus increasing $p_{i\perp}$, the developing ion anisotropy drives the mirror instability to large amplitudes. Saturation in this case is due to ion trapping in the large fluctuations produced by the instability.

3.5.2 An Implicit Hybrid Model

A second new hybrid approach has been investigated by Stanier et al. [69]. This algorithm is unique here because, unlike all of the previous hybrid codes we have discussed in this chapter, it is fully implicit and is designed so that energy and momentum can be conserved to any desired degree of accuracy. (Other hybrid codes do not conserve these due to spatial or temporal truncation errors.) The work builds on previous implicit PIC methods [70, 71] that give energy (but not momentum) conservation. As in other hybrid codes, the magnetic field is advanced in time using Faraday's law, and the electric field is evaluated from the electron momentum equation with $m_e = 0$, i.e., as in Eq. (3.10). It should be noted that in their paper Stanier et al. prefer to use the vector potential \mathbf{A} rather than the magnetic field \mathbf{B} , and they solve a separate equation for the electron pressure that contains terms for resistive heating and electron heat flux. For simplicity of our presentation we will continue here to use \mathbf{B} and assume $p_e = 0$ as well as neglect resistivity. And in the spirit of the previous discussion of other hybrid codes, we will also concentrate on the temporal advance of the electric and magnetic fields.

Stanier et al. [69] assume at time level N , the particle positions \mathbf{x}_p^N and velocities \mathbf{v}_p^N , as well as the magnetic field \mathbf{B}^N are known, while the electric field is known at the previous half-time step, $\mathbf{E}^{N-1/2}$. To advance the electric field to the next time level (always the problem in hybrid codes!), one assumes a guess for the advanced electric field, $\mathbf{E}'^{N+1/2}$. (How this can be done will be discussed later.) With this assumed electric field, the magnetic field can be advanced to time level $N+1$ using Faraday's law to obtain \mathbf{B}'^{N+1} and an equation of the form (Eq. 3.20) can be used to compute $\mathbf{B}''^{N+1/2}$. With these fields at time $N+1/2$, the particles are moved a half-time step ($\Delta t/2$) to $\mathbf{x}_p'^{N+1/2}$ and $\mathbf{v}_p'^{N+1/2}$ where the moments $n_i'^{N+1/2}$ and $\mathbf{V}_i'^{N+1/2}$ are gathered at the same time level. At this stage, the initial guess of $\mathbf{E}'^{N+1/2}$ is

compared with the right-hand side of Ohm's law using Eq. (3.10) (and setting $p_e = 0$ here). The difference between the two (\mathbf{F}) values is called the “residual” and given by

$$\mathbf{F}(\mathbf{E}^{N+1/2}) = \mathbf{E}'^{N+1/2} + \mathbf{V}_i'^{N+1/2} \times \mathbf{B}'^{N+1/2} + \frac{\mathbf{B}'^{N+1/2} \times (\nabla \times \mathbf{B}'^{N+1/2})}{\mu_0 q_i n_i'^{N+1/2}}. \quad (3.22)$$

The initial guess $\mathbf{E}'^{N+1/2}$ is therefore exact when $\mathbf{F}(\mathbf{E}'^{N+1/2}) = 0$, but unfortunately this is extremely unlikely! Instead, the error is calculated as the Euclidean norm, $|\mathbf{F}|$, and compared to a user-specified tolerance, ε . If $|\mathbf{F}| > \varepsilon$, the whole process is repeated with a new guess $\mathbf{E}''^{N+1/2}$ (and potentially $\mathbf{E}'''^{N+1/2}$ and so on) until the condition $|\mathbf{F}| < \varepsilon$ is met and the iteration is stopped. At this point, with $\mathbf{E}'^{N+1/2}$ known, \mathbf{B}^{N+1} is calculated (as well as $\mathbf{B}'^{N+1/2}$), and particle positions \mathbf{x}_p^{N+1} and \mathbf{v}_p^{N+1} are calculated for the next time step using the full time step (Δt). It is found in Stanier et al. [69] that the momentum and energy conservation errors are compatible to the value chosen for ε . Thus, they can be made to be as small as numerical round-off by choosing small ε , but this typically requires more iterations of the above method.

A natural question, and indeed the whole efficiency of the method, depends on the method to determine the electric field guesses. Stanier et al. [69] use a form of the well-known Newton-Krylov methods [72, 73] to do this. \mathbf{E}^{k+1} (the “($k+1$)-th” guess of $\mathbf{E}^{N+1/2}$) is found by iteratively solving the Newton-linearized equations,

$$\left. \frac{\partial \mathbf{F}}{\partial \mathbf{E}} \right|^k (\mathbf{E}^{k+1} - \mathbf{E}^k) = -\mathbf{F}(\mathbf{E}^k), \quad (3.23)$$

using the Flexible-Generalized Minimum Residual method (FGMRES method [74]). Alert readers of this article would note that an initial guess with $k = 0$ is still needed at each time step. The form of this guess is not so critical due to the iterative nature of the algorithm, and it can be taken to be the electric field at the previous time step (extrapolation).

As well as using implicit techniques to solve for the electric field, Stanier et al. [69] also advance the particles with an adaptive and implicit sub-cycling method (in contrast to sub-cycling the fields in the earlier described explicit codes) and use an orbit-averaging technique to reduce the numerical noise (also used previously in an [electrostatic] hybrid code by Sturdevant et al. [75]). The details are probably beyond the concern of most readers of this chapter—the interested reader is referred to the references in Stanier et al. [69]. We do note that sub-cycling the ions is particularly suited for magnetospheric simulations, where ions in the strong dipole field of the Earth have larger gyro-frequencies (and need smaller time steps) than those in the interplanetary magnetic field.

The utility of this method has been analyzed in a number of test problems in Stanier et al. [69]. In addition to several electrostatic problems, the well-known one-dimensional electromagnetic ion cyclotron instability driven by a temperature anisotropy ($T_{\perp}/T_{||} > 1$) (e.g., [76]) that generates the most unstable waves parallel to the background magnetic field and the equally popular GEM reconnection challenge problem [77] have been successfully simulated. Moreover, excellent results have been obtained for the cold ion beam propagation problem, which has been a persistent issue in hybrid codes for some time [78].

In the long run, how well an implicit hybrid code will perform on more complex problems remains to be demonstrated. The strength of implicit methods, with the possibility of using larger time steps, having better control of energy conservation, and suppressing the impact of short-wavelength whistler waves, makes this technique well worth more study. For hybrid codes, and indeed for many other computational plasma physics techniques as well, there is always a trade-off between efficiency, stability, and overall energy conservation issues.

3.5.3 A Brief Note on Comparison of Hybrid Algorithms

Having discussed a number of different hybrid algorithms that have been implemented in various production codes used by a number of researchers on a wide range of problems, we conclude this section with a brief discussion of test problems, numerical checks, and comparisons with different algorithms (since there will always be some differences in the results, as discussed earlier [8]). This discussion is based on our previous tutorial [1] and comparison of results from different hybrid algorithms in the Appendix of Karimabadi et al. [12].

Our earlier tutorial [1] employs two-dimensional simulations based on either the velocity-extrapolation algorithm discussed in Sect. 3.2 or a predictor–corrector formulation described in that article as well as earlier in this section to compare and contrast results of two equal-density, field-aligned counter-streaming ($\pm V_0 = v_A$ = Alfvén speed) ion beams that generate Alfvén waves. Qualitative comparisons of results from runs made with these two different algorithms are described (in their Figs. 1 and 2), along with examining the effects of smoothing (in their Fig. 3) and changing the resistivity (in their Fig. 4) on the spectrum of high-frequency whistlers, ion heating, and energy conservation. The effects on the results obtained in both codes are most noticeable at short wavelengths.

Karimabadi et al. [12] carried out the most detailed and quantitative comparison of the velocity extrapolation, predictor–corrector, and the CAM-CL algorithms in two-dimensional test problems involving a moving discontinuity of two plasma states with differing temperatures and densities and parallel or anti-parallel magnetic fields in the two adjacent regions. The temperatures of the cold ions in each case are compared in the three algorithms and with different size time steps in their Tables A1 and A2. A quick rule of thumb is that all three algorithms give “excellent” results for most practical situations, but in problems where more precision in the wave properties at short wavelength or better energy conservation is needed, not

surprisingly, the more accurate, predictor-corrector algorithm performs better. The presence of short-wavelength whistlers can be reduced by sub-cycling both the electric and magnetic field solves. Some adjustments to these basic algorithms that involve when the ion charge is collected (i.e., at time level $N+1$ at the new particle positions or at time level $N+1/2$ when the velocities have been advanced) that can improve performance in certain problems, such as in cold ion beam propagation [12].

3.6 The Future of Hybrid Codes

We conclude with a brief look into the future using hybrid codes to investigate space plasma phenomena, focusing on three issues. First, there is expected to be continued major community interest in modeling and understanding plasma physics phenomena and their relation to spacecraft observations. This will naturally involve vigorous investigation of processes occurring in the near-Earth environment, e.g., the Earth's bow shock, foreshock, magnetosheath, magnetopause, magnetotail, and observations from MMS and other spacecraft. There will also be sustained interest in phenomena observed at the moon and planets, such as Mercury, Venus, Mars, Jupiter, Saturn, and the other outer planets as well. Processes occurring at the sun and also in the solar wind, as well as at objects residing in the wind, such as comets and asteroids, will also remain of significant importance, especially with the launch of Parker Solar Probe and Solar Orbiter. In all of these situations, as we have seen throughout this chapter, hybrid calculations will play a unique and important role in the interpretation of observations, direct comparison with measurements, and in understanding integrated physical processes (as demonstrated in Sect. 3.3). Overall, these interests will remain the principal driver for hybrid simulations in the foreseeable future.

Second, there will be continuing improvements in algorithms. Especially for hybrid algorithms, one expects continuing efforts to find better ways to reduce effects associated with short-wavelength whistler waves, dynamic methods to improve computational efficiency by varying time steps and/or cell-sizes during the simulation. In addition, in recent years there has been further development of finite-electron mass hybrid models and codes (see Chap. 9), quasi-neutral models (see Chap. 10), implicit methods (see Chap. 8), gyro- and fully kinetic Vlasov as well as PIC methods and codes (see Chaps. 4, 5, and 6, respectively). There is also a permanent interest in developing a local embedding of kinetic electron physics [1]. Such further developments of methods and codes remain challenging, and perhaps important and compelling as well.

Finally, there is ongoing and significantly continued development of new computer architectures, such as exascale computing, of visualization techniques, of asynchronous time-stepping methods (see, e.g., Chap. 13), and of diagnostics in all areas of computational physics. All of these areas are changing very rapidly. Major drivers in these areas are likely to occur in fields other than space physics, where abundant resources are available to bear on their problems. As happened frequently

in the past, space plasma physics will then benefit from these advances in other fields.

Comparing with an earlier (2003) version of this tutorial [1] one can see how computational space plasma physics in the narrow perspective of hybrid codes with mass less electrons has, on the one hand, remained fairly steady. And yet, at the same time, it shows significant changes. The “grand challenge” presented to graduates of the International Space Simulation School ISSS 20 years ago was the issue of embedding kinetic electron physics in hybrid codes. The development of very large-scale PIC codes and the astonishing detailed measurements from MMS spacecraft have rendered this challenge to be of secondary importance. It has now been replaced with the greater challenge of using hybrid codes (3D, of course, but also 2D and even 1D where appropriate) to analyze and interpret the complex integrated plasma physics that occurs at all boundary layers in space and problems of plasma turbulence (see Chap. 9).

References

1. D. Winske, L. Yin, N. Omidi, H. Karimabadi, K. Quest, Hybrid simulation codes: past, present and future, in *Space Plasma Simulation*, ed. by J. Büchner, C. T. Dum, M. Scholer, (Springer, Berlin, 2003), pp. 136–165
2. J.L. Burch, T.E. Moore, R.B. Torbert, B.L. Giles, Magnetospheric multiscale overview and science objectives. *Space Sci. Rev.* **199**, 5–21 (2016). <https://doi.org/10.1007/s11214-015-0164-9>
3. E.W. Greenstadt, C.T. Russell, J.T. Gosling, G. Paschmann, G.K. Parks, K. Anderson, F.L. Scarf, et al., A macroscopic profile of the typical quasi-perpendicular bow shock: ISEE-1 and 2. *J. Geophys. Res.* **85**, 2124–2130 (1980). <https://doi.org/10.1029/JA085iA05p02124>
4. J.L. Burch, R.B. Torbert, T.D. Phan, L.-J. Chen, T.E. Moore, R.E. Ergun, J.P. Eastwood, et al., Electron scale measurements of magnetic reconnection in space. *Science* **352**(6290) aaf2939, 1–10 (2016). <https://doi.org/10.1126/science.aaf.239>
5. C.K. Birdsall, A.B. Langdon, *Plasma Physics via Computer Simulation*, Ch. 8 (Institute of Physics Publishing, Bristol UK, 1991)
6. C.W. Nielson, H.R. Lewis, Particle code methods in the non-radiative limit, in *Methods in Computational Physics*, ed. by J. Killeen, vol. 16, (Academic Press, New York, 1976), pp. 367–388
7. R.L. Morse, C.W. Nielson, Numerical simulation of the Weibel instability in one and two dimensions. *Phys. Fluids* **14**, 830–840 (1971). <https://doi.org/10.1063/1.1693518>
8. A. Stanier, L. Chacón, A. Le, A cancellation problem in hybrid particle-in-cell schemes due to finite particle size. *J. Comp. Phys.* **420**, 109705 (2020). <https://doi.org/10.1016/j.jcp.2020109705>
9. 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 Scientific, Tokyo, 1993), pp. 103–160
10. D.S. Harned, Quasineutral hybrid simulation of macroscopic plasma phenomena. *J. Comp. Phys.* **47**, 452–462 (1982). [https://doi.org/10.1016/0021-9991\(82\)90094-8](https://doi.org/10.1016/0021-9991(82)90094-8)
11. A.P. Matthews, Current advance method and cyclic leapfrog for 2D multispecies hybrid plasma simulations. *J. Comp. Phys.* **112**, 102–116 (1994). <https://doi.org/10.1006/jcph.1994.1084>
12. H. Karimabadi, D. Krauss-Varban, J.D. Huba, H.X. Vu, On magnetic reconnection regimes and associated three-dimensional asymmetries: hybrid, Hall-less hybrid and Hall-MHD simulations. *J. Geophys. Res.* **109**, A09205 (2004). <https://doi.org/10.1029/2004JA010478>

13. T. Terasawa, M. Hoshino, J.I. Sakai, T. Hada, Decay instability of finite amplitude circularly polarized Alfvén waves: a numerical simulation of stimulated Brillouin scattering. *J. Geophys. Res.* **91**, 4171–4187 (1986). <https://doi.org/10.1029/J091iA04p04171>
14. N. Omidi, Formation of foreshock cavities, in *Turbulence and Nonlinear Processes in Astrophysical Plasmas. AIP Conference Proceedings*, ed. by D. Shaikh, G. Zank, vol. 932, (Amer. Inst. Phys., New York, 2007), p. 181
15. X. Blanco-Cano, N. Omidi, C.T. Russell, Global hybrid simulations: foreshock waves and cavities under radial IMF geometry. *J. Geophys. Res.* **114**, A01216 (2009). <https://doi.org/10.1029/2008JA013406>
16. P. Kajdič, X. Blanco-Cano, N. Omidi, C.T. Russell, Multispacecraft study of foreshock cavities. *Planet. Space Sci.* **59**, 705–714 (2011). <https://doi.org/10.1016/j.pss.2011.02.005>
17. N. Omidi, G. Collinson, D. Sibeck, Structure and properties of the foreshock at Venus. *J. Geophys. Res. Space Phys.* **122**, 10275–10286 (2017). <https://doi.org/10.1002/2017JA024180>
18. N. Omidi, D. Sibeck, X. Blanco-Cano, The foreshock compressional boundary. *J. Geophys. Res.* **114**, A08205 (2009). <https://doi.org/10.1029/2008JA01395>
19. N. Omidi, G. Collinson, D. Sibeck, Foreshock bubbles at Venus: hybrid simulations and VEX observations. *J. Geophys. Res. Space Phys.* **125**, e2019JA027056 (2020). <https://doi.org/10.1029/2019JA027056>
20. N. Omidi, H. Zhang, D. Sibeck, D. Turner, Spontaneous hot flow anomalies at quasi-parallel shocks: 2. Hybrid simulations. *J. Geophys. Res.* **118**, 173–180 (2013). <https://doi.org/10.1029/2012JA018099>
21. J. Egedal, A. Le, W. Daughton, A review of pressure anisotropy caused by electron trapping in collisionless plasma and its implications for magnetic reconnection. *Phys. Plasmas* **20**, 061201 (2013). <https://doi.org/10.1063/1.4811092>
22. J. Egedal, M. Øieroset, W. Fox, R.P. Lin, In situ discovery of an electrostatic potential, trapping electrons and mediating fast reconnection in the Earth's magnetotail. *Phys. Rev. Lett.* **94**, 025006 (2005). <https://doi.org/10.1103/PhysRevLett.94.025006>
23. A. Le, J. Egedal, W. Daughton, W. Fox, N. Katz, Equations of state for collisionless guide-field reconnection. *Phys. Rev. Lett.* **102**, 085001 (2009). <https://doi.org/10.1103/PhysRevLett.102.085001>
24. A. Le, W. Daughton, H. Karimabadi, J. Egedal, Hybrid simulations of magnetic reconnection with kinetic ions and fluid electron pressure anisotropy. *Phys. Plasmas* **23**, 032114 (2016). <https://doi.org/10.1063/1.4943893>
25. M. Øieroset, T.D. Phan, C. Haggerty, M.A. Shay, J.P. Eastwood, D.J. Gershman, et al., MMS observations of large guide field symmetric reconnection between colliding reconnection jets at the center of a magnetic flux rope at the magnetopause. *Geophys. Res. Lett.* **43**, 5536–5544 (2016). <https://doi.org/10.1002/2016GL069166>
26. A. Le, J. Egedal, O. Ohia, W. Daughton, H. Karimabadi, V.S. Lukin, Regimes of the electron diffusion region in magnetic reconnection. *Phys. Rev. Lett.* **110**, 135004 (2013). <https://doi.org/10.1103/PhysRevLett.110.135004025006>
27. B.A. Wetherton, J. Egedal, A. Le, W. Daughton, Validation of anisotropic electron fluid closure through in situ spacecraft observations of magnetic reconnection. *Geophys. Rev. Lett.* **46**, 6223–6229 (2020). <https://doi.org/10.1002/2020GL069166>
28. H. Karimabadi, V. Roytershten, H.X. Vu, Y.A. Omelchenko, J. Scudder, W. Daughton, A. Dimmock, K. Nyjyri, M. Wan, D. Sibeck, M. Tatineni, A. Majumdar, B. Loring, B. Geveci, The link between shocks, turbulence and magnetic reconnection in collisionless plasmas. *Phys. Plasmas* **21**, 062308 (2014). <https://doi.org/10.1063/1.4882875>
29. H. Karimabadi, B. Loring, H.X. Vu, Y. Omelchenko, M. Tatineni, A. Majumdar, U. Ayachit, B. Geveci, in *Petascale Global Kinetic Simulations of the Magnetosphere and Visualization Strategies for Analysis of Very Large Multivariate Data Sets*. 5th International Conference of Numerical Modeling of Space Plasma Flows, Astronom (2010)
30. A. Retinò, D. Sundkvist, A. Vaivads, F. Mozer, M. André, C.J. Owen, In situ evidence of magnetic reconnection in turbulent plasma. *Nat. Phys.* **3**, 236–238 (2007). <https://doi.org/10.1038/nphys5574>

31. Z. Vörös, E. Yordanova, A. Varsani, K.J. Genestreti, Y.V. Khotyaintsev, W. Li, et al., MMS observation of magnetic reconnection in the turbulent magnetosheath. *J. Geophys. Res.* **122**, 11442–11467 (2017). <https://doi.org/10.1002/2017JA024535>
32. E. Eriksson, A. Vaivads, D.B. Graham, A. Divin, Y.V. Khotyaintsev, E. Yordanova, et al., Electron energization at a reconnecting magnetosheath current sheet. *Geophys. Res. Lett.* **45**, 8081–8090 (2018). <https://doi.org/10.1029/2018GL078660>
33. P.C. Liewer, M. Velli, B.E. Goldstein, Alfvén wave propagation and ion cyclotron interactions in the expanding solar wind: one-dimensional hybrid simulations. *J. Geophys. Res.* **106**, 29261–29281 (2001). <https://doi.org/10.1029/2001JA000086>
34. P. Hellinger, P.M. Trávníček, A. Mangeney, R. Grappin, Hybrid simulations of the expanding solar wind: temperature and drift velocities. *Geophys. Res. Lett.* **30**, 1211–1214 (2003). <https://doi.org/10.1029/2002GL016409-2003>
35. P. Hellinger, P.M. Trávníček, Oblique proton fire hose instability in the expanding solar wind: hybrid simulations. *J. Comp. Phys.* **113**, A10109 (2008). <https://doi.org/10.1029/2008JA013416>
36. P. Hellinger, L. Matteini, S. Landi, A. Verdini, L. Franci, P.M. Trávníček, Plasma turbulence and kinetic instabilities at ion scales in the expanding solar wind. *Ap. J. Lett.* **811**, L32 (2015). <https://doi.org/10.1088/2041-8205/811/2/L32>
37. P. Hellinger, S. Landi, L. Matteini, A. Verdini, L. Franci, Mirror instability in the turbulent solar wind. *Ap. J.* **838**, 158 (2017). <https://doi.org/10.3847/1538-4357/aa67e0>
38. L. Franci, S. Landi, L. Matteini, A. Verdini, P. Hellinger, High-resolution hybrid simulations of kinetic plasma turbulence at proton scales. *Ap. J.* **812**, 21 (2015). <https://doi.org/10.1088/0004-637X/812/1/21>
39. L. Franci, S. Landi, L. Matteini, A. Verdini, P. Hellinger, Plasma beta dependence of the ion-scale spectral break of solar wind turbulence: high-resolution 2D hybrid simulations. *Ap. J.* **833**, 91 (2016). <https://doi.org/10.3847/1538-4357/833/1/91>
40. L. Franci, S. Landi, A. Verdini, L. Matteini, P. Hellinger, Solar wind turbulence cascade from MHD to sub-ion scales: large-size 3D hybrid particle-in-cell simulations. *Ap. J.* **853**, 26 (2018). <https://doi.org/10.3847/1538-4357/aaa3e8>
41. S.H. Brecht, S.A. Ledvina, B.M. Jakosky, The role of the electron temperature on ion loss from Mars. *J. Geophys. Res.* **122**, 8375–8390 (2017). <https://doi.org/10.1002/2016JA023510>
42. B.M. Jakosky, J.M. Grebowsky, J.G. Luhmann, D.A. Brain, Initial results from the MAVEN mission to Mars. *Geophys. Res. Lett.* **42**, 8791–8802 (2015). <https://doi.org/10.1002/2015GL065271>
43. S.A. Ledvina, S.H. Brecht, D.A. Brain, B.M. Jakosky, Ion escape rates from Mars: results from hybrid simulations compared to MAVEN observations. *J. Geophys. Res.* **122**, 8391–8408 (2017). <https://doi.org/10.1002/2016JA023521>
44. A.S. Lipatov, J.F. Cooper, E.C. Sittler, R.E. Hartle, The light (H^+ , H_2^+ , He^+) and heavy (Na^+) pickup ion dynamics in the lunar-like plasma environment: 3D hybrid kinetic modeling. *Adv. Space Res.* **52**, 1929–1938 (2013). <https://doi.org/10.1016/j.asr.2013.08.023>
45. S. Wiehle, F. Plaschke, U. Motschmann, K.-H. Glassmeier, H.U. Auster, V. Angelopoulos, J. Mueller, H. Kriegel, E. Georgescu, J. Halekas, D.G. Sibeck, J.P. McFadden, First lunar wake passage of ARTEMIS: discrimination of wake effects and solar wind fluctuations by 3D hybrid simulations. *Planet. Space Sci.* **59**, 661–671 (2011). <https://doi.org/10.1016/j.pss.2011.01.012>
46. P.M. Trávníček, P. Hellinger, D. Schriver, D. Hercik, J.A. Slavin, B.J. Anderson, Kinetic instabilities in Mercury's magnetosphere: three-dimensional simulation results. *Geophys. Res. Lett.* **36**, L07104 (2009). <https://doi.org/10.1029/2008GL036630>
47. B.J. Anderson, M.H. Acuna, H. Korth, M.E. Purucker, C.L. Johnson, J.A. Slavin, S.C. Solomon, R.L. McNutt, The structure of Mercury's magnetic field from MESSENGER's first flyby. *Science* **321**, 82–85 (2008). <https://doi.org/10.1126/science.1159081>
48. P.M. Trávníček, D. Schriver, P. Hellinger, D. Hercik, B.J. Anderson, M. Sarantos, J.A. Slavin, Mercury's magnetosphere-solar wind interaction for northward and southward interplanetary magnetic field: hybrid simulation results. *Icarus* **209**, 11–22 (2010). <https://doi.org/10.1016/j.icarus.2010.01.008>

49. D. Burgess, P. Hellinger, I. Gingell, P.M. Trávníček, Microstructure in two- and three-dimensional hybrid simulations of perpendicularly collisionless shocks. *J. Plasma Phys.* **82**, 1–23 (2016). <https://doi.org/10.1017/S0022377816000660>
50. L. Gargaté, R.A. Fonseca, L.O. Silva, R.A. Bamford, R. Bingham, SEP acceleration in CME driven shocks using a hybrid code. *Ap. J.* **792**, 9 (2014). <https://doi.org/10.1088/0004-637X/792/1/9>
51. L. Gargaté, R. Bingham, R.A. Fonseca, L.O. Silva, dHybrid: a massively parallel code for hybrid simulations of space plasmas. *Comp. Phys. Comm.* **176**, 419–425 (2007). <https://doi.org/10.1016/j.cpc.2006.11.013>
52. D. Caprioli, A. Spitkovsky, Simulations of ion acceleration at non-relativistic shocks. I. Acceleration efficiency. *Ap. J.* **783**, 91 (2014). <https://doi.org/10.1088/0004-637X/783/2/91>
53. L. Vshivkova, G. Dudnikova, K. Vshivkov, Hybrid numerical model of shock waves in collisionless plasma, in *AIP Conference Proceedings*, vol. 1173, (Amer. Inst. Phys., New York, 2016), p. 110017. <https://doi.org/10.1063/1.4965021>
54. F. Guo, J. Giacalone, The acceleration of thermal protons at parallel collisionless shocks: three-dimensional hybrid simulations. *Ap. J.* **773**, 158 (2013). <https://doi.org/10.1088/0004-637X/773/2/158>
55. Y. Lin, J.R. Johnson, X. Wang, Three-dimensional mode conversion associated with kinetic Alfvén waves. *Phys. Rev. Lett.* **109**, 125003 (2012). <https://doi.org/10.1103/PhysRevLett.109.125003>
56. D.W. Swift, Use of a hybrid code for a global-scale plasma simulation. *J. Comp. Phys.* **126**, 109–121 (1996). <https://doi.org/10.1006/jcph.1996.0124>
57. Z. Guo, M. Hong, Y. Lin, A. Du, X. Wang, M. Wu, Q. Lu, Generation of kinetic Alfvén waves in the high-latitude near-Earth magnetotail: a global hybrid simulation. *Phys. Plasmas* **22**, 022117 (2015). <https://doi.org/10.1063/1.4907666>
58. D.W. Hewett, C.W. Nielson, A multidimensional quasineutral plasma simulation model. *J. Comp. Phys.* **29**, 219–236 (1978). [https://doi.org/10.1016/0021-9991\(78\)90153-5](https://doi.org/10.1016/0021-9991(78)90153-5)
59. M.A. Shay, J.F. Drake, R.E. Denton, D. Biskamp, Structure of the dissipation region during collisionless magnetic reconnection. *J. Geophys. Res.* **103**, 9165–9176 (1998). <https://doi.org/10.1029/97JA03528>
60. A. Le, W. Daughton, O. Ohia, L.-J. Chen, Y.-H. Liu, S. Wang, W.D. Nystrom, R. Bird, Drift turbulence, particle transport and anomalous dissipation at the reconnecting magnetopause. *Phys. Plasmas* **25**, 062103 (2018). <https://doi.org/10.1063/1.5027086>
61. A.S. Lipatov, *The Hybrid Multi-Scale Simulation Technology* (Springer, Berlin, 2002)
62. T. Amano, A generalized quasi-neutral fluid-particle hybrid plasma model and its application to energetic-particle-magnetohydrodynamics hybrid simulation. *J. Comp. Phys.* **366**, 366–285 (2018). <https://doi.org/10.1016/j.jcp.2018.04.020>
63. P.A. Muñoz, N. Jain, P. Kilian, J. Büchner, A new hybrid code (CHIEF) implementing the inertial electron fluid equation without approximation. *Comp. Phys. Comm.* **224**, 245–264 (2018). <https://doi.org/10.1016/j.cpc.2017.10.012>
64. G. Lapenta, J.U. Brackbill, P. Ricci, Kinetic approach to microscopic-macroscopic coupling in space and laboratory plasmas. *Phys. Plasmas* **13**, 055904 (2006). <https://doi.org/10.1063/1.2173623>
65. K. Noguchi, C. Tronci, G. Zuccaro, G. Lapenta, Formulation of the relativistic moment implicit particle-in-cell method. *Phys. Plasmas* **14**, 042308 (2007). <https://doi.org/10.1063/1.2721083>
66. Y.A. Omelchenko, H. Karimabadi, HYPERs: a unidimensional asynchronous framework for multiscale hybrid simulations. *J. Comp. Phys.* **231**, 1766–1780 (2012). <https://doi.org/10.1016/j.jcp.2011.11.004>
67. M.W. Kunz, J.M. Stone, X.-N. Bai, Pegasus: a new hybrid-kinetic particle-in-cell code for astrophysical particle dynamics. *J. Comp. Phys.* **259**, 154–174 (2014). <https://doi.org/10.1016/j.jcp.2013.11.0035>
68. M.W. Kunz, A.A. Schekochihin, J.M. Stone, Firehose and mirror instabilities in a collisionless shearing plasma. *Phys. Rev. Lett.* **112**, 205003 (2014). <https://doi.org/10.1103/PhysRevLett.112.205003>

69. A. Stanier, L. Chacón, G. Chen, A fully implicit conservative nonlinear electromagnetic hybrid particle ion/fluid electron algorithm. *J. Comp. Phys.* **376**, 597–616 (2019). <https://doi.org/10.1016/j.jcp.2018.09.038>
70. G. Chen, L. Chacón, D.C. Barnes, An energy- and charge-conserving implicit electrostatic particle-in-cell algorithm. *J. Comp. Phys.* **230**, 7018–7036 (2011). <https://doi.org/10.1026/j.jcp.2011.05.031>
71. S. Markidis, G. Lapenta, The energy conserving particle-in-cell method. *J. Comp. Phys.* **230**, 7037–7052 (2011). <https://doi.org/10.1016/j.jcp.2011.05.033>
72. C.T. Kelley, *Iterative Methods for Linear and Nonlinear Equations* (SIAM, Philadelphia, 1995)
73. D.A. Knoll, D.E. Keyes, Jacobian-free Newton-Krylov methods: a survey of approaches and applications. *J. Comp. Phys.* **193**, 357–397 (2004). <https://doi.org/10.1016/j.jcp.2003.08.010>
74. Y. Saad, A flexible inner-outer preconditioned gmres algorithm. *SIAM J. Sci. Comp.* **14**, 461–469 (1993). <https://doi.org/10.1137/0914028>
75. B.J. Sturdevant, S.E. Parker, Y. Chen, B.B. Hause, An implicit δf particle-in-cell method with sub-cycling and orbit averaging for Lorentz ions. *J. Comp. Phys.* **316**, 519–533 (2016). <https://doi.org/10.1016/j.jcp.2016.04.036>
76. S.P. Gary, S.A. Fuselier, B.J. Anderson, Ion anisotropy instabilities in the magnetosheath. *J. Geophys. Res.* **98**, 1481–1488 (1993). <https://doi.org/10.1029/92JA01844>
77. 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, Geospace environmental modeling (GEM) magnetic reconnection challenge. *J. Geophys. Res.* **106**, 3715–3720 (2001). <https://doi.org/10.1029/1999JA900449>
78. P.W. Rambo, Finite-grid instability in quasi-neutral hybrid simulations. *J. Comp. Phys.* **118**, 152–158 (1995). <https://doi.org/10.1006/jcph.1995.1086>



Gyrokinetics

4

Frank Jenko

Abstract

Over the last several years, gyrokinetics has emerged as a new approach to studying magnetized, weakly collisional space plasmas, complementing fully kinetic and hybrid kinetic-fluid approaches. Initial applications include, in particular, plasma turbulence at kinetic scales and magnetic reconnection in the presence of guide fields. The present chapter consists of an accessible introduction to gyrokinetics, an overview of computational aspects of gyrokinetics, and a summary of key results obtained by applying gyrokinetics to turbulent heating in the solar wind. It ends with a brief outlook regarding future work.

4.1 Effective Kinetic Description of Magnetized, Weakly Collisional Space Plasmas

For many years, computational space plasma physics has been employing kinetic descriptions to investigate weakly collisional systems. This is often necessary to capture important kinetic phenomena like Landau damping, finite Larmor radius effects, or ion/electron cyclotron resonances—which are not retained in conventional fluid descriptions. Here, each particle species is represented by a time-dependent single-particle distribution function $f(\mathbf{x}, \mathbf{v}, t)$ in position-velocity space (“phase space”), which may exhibit significant deviations from a Maxwell-Boltzmann distribution. By calculating velocity space integrals (“moments”), one obtains the density, mean velocity, temperature, heat flux, etc. of the respective particle species. Technically, a kinetic description is equivalent to an infinite hier-

F. Jenko (✉)

Max Planck Institute for Plasma Physics, Garching, Germany
e-mail: frank.jenko@ipp.mpg.de

archy of coupled fluid equations involving these moments. For practical purposes, this hierarchy needs to be truncated, however, and this is where kinetic effects are typically lost.

4.1.1 Limits of Fully Kinetic and Hybrid Kinetic-Fluid Descriptions

Arguably, the most direct way to implement a kinetic description is based on the fully kinetic Vlasov-Maxwell system of equations. Here, the equation of motion for each $f_s(\mathbf{x}, \mathbf{v}, t)$ is given by¹

$$\frac{\partial f_s}{\partial t} + \mathbf{v} \cdot \frac{\partial f_s}{\partial \mathbf{x}} + \frac{q_s}{m_s} \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right) \cdot \frac{\partial f_s}{\partial \mathbf{v}} = 0, \quad (4.1)$$

where the subscript s denotes the species label; q_s and m_s are, respectively, the species' charge and mass; and c is the speed of light. The self-consistent electric and magnetic fields, \mathbf{E} and \mathbf{B} , are then calculated via Maxwell's equations, using the densities

$$n_s(\mathbf{x}, t) \equiv \int f_s(\mathbf{x}, \mathbf{v}, t) d^3 v \quad (4.2)$$

and mean velocities

$$\mathbf{u}_s(\mathbf{x}, t) \equiv \frac{1}{n_s(\mathbf{x}, t)} \int \mathbf{v} f_s(\mathbf{x}, \mathbf{v}, t) d^3 v \quad (4.3)$$

to determine the charge and current densities. For more details on fully kinetic equations and numerical techniques to solve them, the reader is referred to Chaps. 5 and 6 of this book.

The Vlasov-Maxwell system contains a large variety of plasma phenomena down to very small spatiotemporal scales. It captures, for example, Debye shielding, Langmuir waves, cyclotron resonances, and even electromagnetic waves. This is an important feature if one is explicitly interested in such small-scale dynamics. However, often one is not, and numerical solutions of the Vlasov-Maxwell system in six-dimensional phase space can be difficult or even impossible to get for realistic physical parameters and system sizes, even on the world's most powerful supercomputers. Therefore, reduced kinetic descriptions are often used in practice.

One popular example for carrying out such a model reduction is to replace the kinetic description of the electrons by a relatively simple fluid model while retaining a kinetic description of the ions. Such a hybrid kinetic-fluid approach (see Chapter 3 of this book) eliminates many complications of the fully kinetic model and leads to a much more tractable equations. However, kinetic effects related to the electron

¹ Here and throughout this chapter, we employ cgs units.

dynamics are removed from the system. This includes, in particular, electron Landau damping which affects, for example, the properties of Alfvén waves (even at long wavelengths) and acts as an important catalyst for energy dissipation in weakly collisional plasmas (as will be discussed in more details in Sect. 4.4).

While there have been attempts to eliminate this problem by resorting to the so-called gyrofluid electron models (see, for example, [1–3]), which contain extra terms to mimic electron Landau damping, high-fidelity realizations of this idea are difficult to implement [4–6]. One is therefore left wondering if there are other ways to reduce the fully kinetic model, complementing the well-established hybrid kinetic-fluid approach. As it turns out, gyrokinetics provides a perfect opportunity to fill this gap.

4.1.2 The New Kid on the Block: Gyrokinetics

Beginning in the late 1960s, with several breakthrough results emerging throughout the 1980s, theoretical and computational physicists in magnetic confinement fusion research aimed to develop an effective kinetic description of magnetized, weakly collisional plasmas [7–9]. One of the main goals was to develop a solid basis for the investigation of small-scale turbulence in hot, dilute fusion plasmas. In this context, it was noted that the observed turbulent fluctuations exhibit a number of interesting characteristics:

- Typical amplitudes of the turbulent fluctuations (in the densities, temperatures, etc.) tend to be small compared to the respective mean values.
- Typical frequencies of the turbulent fluctuations can cover a wide range of scales but are generally small compared to the ion and electron gyrofrequencies.
- Typical perpendicular correlation lengths of the turbulent fluctuations tend to be small compared to various macroscopic scale lengths but may be comparable to the ion and electron Larmor radii.
- Typical parallel correlation lengths of the turbulent fluctuations tend to be comparable to various macroscopic scale lengths.

These observations define a peculiar parameter ordering (which is called “gyrokinetic ordering” and will be discussed in more details in Sect. 4.2.2.1) and can serve as a starting point for the derivation of reduced kinetic equations, namely, those of gyrokinetics.

Given that all the particles’ gyration times about the magnetic field lines are much shorter than any other dynamical time scales, it seems natural to try to eliminate the fast gyromotion from the rest of the dynamics. This can be achieved by focusing on the dynamics of the respective guiding centers and by replacing the helical particle trajectories by rings which move more or less freely along magnetic field lines while drifting slowly perpendicular to them. That allows us to remove the explicit treatment of the gyration while fully retaining many key effects (like those linked to

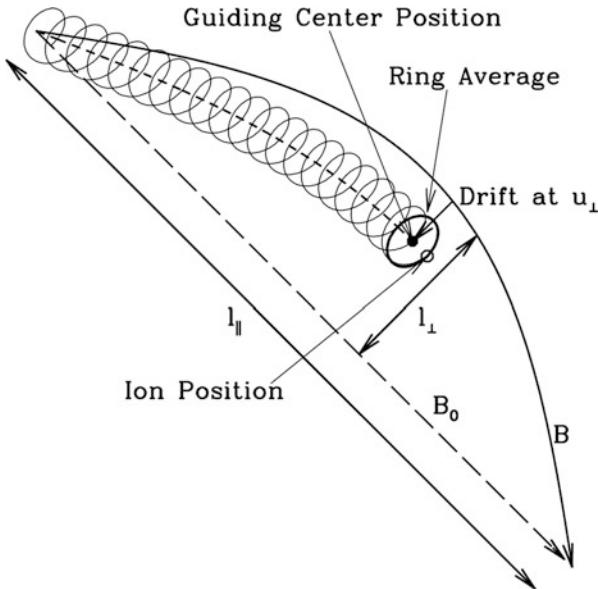


Fig. 4.1 Illustration of the basic idea behind gyrokinetics. Figure from [10] reproduced with permission by AAS. A charged particle embedded in a strong magnetic field exhibits a helical trajectory about the time-varying guiding center position. The motion of the particle is replaced by the motion of the guiding center, with the forces acting on the particle being averaged along respective rings

electron Landau damping or to the existence of finite Larmor radii). This basic idea is illustrated in Fig. 4.1.

Such a reduction has two consequences: First, the fast dynamics of the gyrophase is completely removed (mathematically speaking, averaged out or decoupled), leaving only two velocity space coordinates and thus reducing the kinetic problem from six to five dimensions. This leads to considerable computational savings in and of itself. Second, and much more importantly, various phenomena on small spatiotemporal scales (including, in particular, Debye shielding and high-frequency waves, from electromagnetic waves to Langmuir waves) are eliminated, allowing for *computational savings of many orders of magnitude*. This is what makes gyrokinetics so very attractive, provided the above conditions are satisfied. As it turns out, a fair number of important space physics problems fall into that category. In recent years, this realization has led to a constantly growing effort to apply the gyrokinetic approach to space plasmas.

4.1.3 Comparing Gyrokinetic, Hybrid Kinetic-Fluid, and Fully Kinetic Descriptions: Waves and Turbulence in Space Plasmas

In order to assess the relative strengths and weaknesses of the gyrokinetic, hybrid kinetic-fluid, and fully kinetic descriptions, it is instructive to compare them directly. Below, two examples of such studies are given, focused, respectively, on waves and turbulence in space plasmas.

In Ref. [11], the linear physics of kinetic Alfvén waves was investigated systematically, focusing on physical parameters that are representative of the solar wind at about 1 AU. It was found that gyrokinetics accurately captures the properties of kinetic Alfvén waves as described by the fully kinetic model, as long as the relevant frequencies stay below the ion cyclotron frequency. Meanwhile, the hybrid approach underpredicts the linear wave damping rates even on ion kinetic scales across a wide parameter range, due to the lack of electron Landau damping (see Fig. 4.2). Similar issues arise in relation to other plasma waves. This aspect has far-reaching consequences which are sometimes not fully appreciated when employing the hybrid model to various physics problems.

In Ref. [12], all three types of models were used side by side to study collisionless plasma turbulence at kinetic scales in two spatial dimensions. For an ion plasma beta of $\beta_i = 0.5$, which is typical of the solar wind at about 1 AU, gyrokinetics agrees well with the fully kinetic model (in terms of various spectral properties) all the way from ion to electron kinetic scales. This provides evidence to support the claim that in this parameter regime, the turbulent dynamics tends to be dominated by a kinetic Alfvén wave cascade as captured by gyrokinetics. Meanwhile, the hybrid model exhibits spectral slopes which are less steep than the fully kinetic and gyrokinetic ones, due to the lack of electron Landau damping.

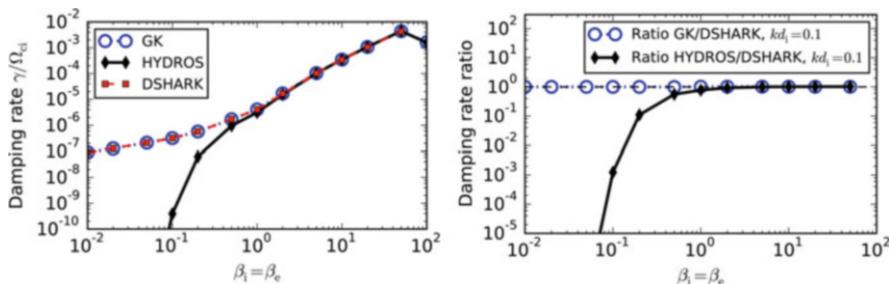


Fig. 4.2 Comparison of damping rates of kinetic Alfvén waves as calculated by fully kinetic (“DSHARK”), hybrid kinetic (“HYDROS”), and gyrokinetic (“GK”) codes as a function of the ion/electron plasma beta. Normalization is with respect to the ion cyclotron frequency. Figures from [11] reprinted with permission by the *New Journal of Physics* (CC-BY 4.0). Once the plasma beta drops below unity, the damping rates are dominated by electron Landau damping, resulting in severely underpredicted damping rates in the hybrid kinetic model. Meanwhile, the gyrokinetic model captures this effect well

These findings illustrate the capabilities and advantages of gyrokinetics within its range of validity.

4.1.4 Applications of Gyrokinetics in Space Plasma Physics

Over the last decade or so, it has become increasingly clear that gyrokinetics can be applied to a large number of space plasma physics problems, shedding light on questions like these:

- What is the nature of space plasma turbulence at kinetic scales?
- How is turbulent energy dissipated in the virtually collisionless solar wind?
- How can turbulent dissipation contribute to coronal heating?
- What is the character of magnetic reconnection in the presence of a guide field?
- How do magnetic reconnection and turbulence interact in magnetized plasmas?

This list could be continued, of course, but suffice it to give some key examples, providing many good reasons to engage with gyrokinetics.

4.2 A Primer on Gyrokinetics

Having described the motivation for the development of gyrokinetics as well as some of the underlying physical ideas, it is time now to address its mathematical formulation. In the context of this introduction to gyrokinetics, the goal will be to establish a solid foundation for newcomers to this field—not to give a full account—which would require much more space than is available here. At the end of this section, there will be recommendations for further reading.

4.2.1 Guiding Center Dynamics in Given Electromagnetic Fields

In principle, it is possible to use the Vlasov-Maxwell equations as a starting point to derive the gyrokinetic equations in the spirit of Frieman and Chen [13]. In a first step, one can write down the equations of motion for charged particles in given electromagnetic fields, transition from particle coordinates to guiding center coordinates, and then average out the fast gyromotion. However, this procedure is somewhat cumbersome and involves a number of fairly subtle points. Special care is required, for instance, to ensure that various fundamental conservation laws are not broken. Given these challenges, it is worth considering an alternative route.

4.2.1.1 A Lagrangian Approach

At this point, one should recall that the original equations of motion under consideration,

$$\dot{\mathbf{x}} = \mathbf{v}, \quad \dot{\mathbf{v}} = \frac{q}{m} \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right), \quad (4.4)$$

can be derived within the framework of Lagrangian mechanics from the Lagrangian

$$\mathcal{L} = \left(\frac{q}{c} \mathbf{A} + m\mathbf{v} \right) \cdot \dot{\mathbf{x}} - \left(\frac{m}{2} v^2 + q\phi \right) \quad (4.5)$$

via the Euler-Lagrange equations,

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}_i} - \frac{\partial \mathcal{L}}{\partial q_i} = 0, \quad (4.6)$$

with $\mathbf{q} = (\mathbf{x}, \mathbf{v})$ as a six-component vector. Here, the vector and scalar potentials, \mathbf{A} and ϕ , and the electric and magnetic fields,

$$\mathbf{E} = -\nabla\phi - \partial_t \mathbf{A}/c, \quad \mathbf{B} = \nabla \times \mathbf{A}, \quad (4.7)$$

are generally dependent on \mathbf{x} and t . As it turns out, it is easier and more transparent to make the transition from particle to guiding center coordinates on the level of the Lagrangian than on the level of the equations of motion. This step is described next.

4.2.1.2 From Particle Coordinates to Guiding Center Coordinates

First, let us transition from particle coordinates (\mathbf{x}, \mathbf{v}) to guiding center coordinates $(\mathbf{X}, v_{\parallel}, \mu, \varphi)$, where \mathbf{X} is the guiding center position, v_{\parallel} is the component of \mathbf{v} parallel to the local magnetic field, $\mu \equiv (mv_{\perp}^2)/(2B)$ is the magnetic moment, and φ is the gyrophase. Here, \mathbf{x} and \mathbf{X} are linked by means of the relationship $\mathbf{x} = \mathbf{X} + \mathbf{r}$, where $\mathbf{r} = \mathbf{v} \times \mathbf{b}/\Omega$ is the vector pointing from the gyrocenter position to the particle position. Here, $\mathbf{b} = \mathbf{B}/B$ is the unit vector along the local magnetic field, and $\Omega = (qB)/(mc)$ is the local gyrofrequency.

Based on this change of coordinates, we will now rewrite each individual term in Eq. (4.5). Using the first-order Taylor expansion $\mathbf{A}(\mathbf{x}) \simeq \mathbf{A}(\mathbf{X}) + (\mathbf{r} \cdot \nabla)\mathbf{A}(\mathbf{X})$, one obtains

$$\mathbf{A}(\mathbf{x}) \cdot \dot{\mathbf{x}} \simeq \mathbf{A}(\mathbf{X}) \cdot \dot{\mathbf{X}} + (\mathbf{r} \cdot \nabla)\mathbf{A}(\mathbf{X}) \cdot \dot{\mathbf{X}} + \mathbf{A}(\mathbf{X}) \cdot \dot{\mathbf{r}} + (\mathbf{r} \cdot \nabla)\mathbf{A}(\mathbf{X}) \cdot \dot{\mathbf{r}}. \quad (4.8)$$

Removing the fast gyrophase dependence by averaging this expression over the gyrophase φ (from 0 to 2π), we get

$$\frac{q}{c} \mathbf{A}(\mathbf{x}) \cdot \dot{\mathbf{x}} \simeq \frac{q}{c} \mathbf{A}(\mathbf{X}) \cdot \dot{\mathbf{X}} - \frac{\mu B}{\Omega} \dot{\varphi}, \quad (4.9)$$

where the second and third term on the right-hand side of Eq. (4.8) drop out upon gyroaveraging, while the fourth term survives. B and Ω are to be evaluated at the guiding center position. Similarly, we get

$$m\mathbf{v} \cdot \dot{\mathbf{x}} = mv_{\parallel}\mathbf{b} \cdot \dot{\mathbf{X}} + mv_{\parallel}\mathbf{b} \cdot \dot{\mathbf{r}} + m\mathbf{v}_{\perp} \cdot \dot{\mathbf{X}} + m\mathbf{v}_{\perp} \cdot \dot{\mathbf{r}}, \quad (4.10)$$

which becomes

$$m\mathbf{v} \cdot \dot{\mathbf{x}} = mv_{\parallel}\mathbf{b} \cdot \dot{\mathbf{X}} + \frac{2\mu B}{\Omega}\dot{\phi}, \quad (4.11)$$

upon gyroaveraging. Here, again, the second and third term on the right-hand side drop out, while the fourth term survives, and \mathbf{b} , B , and Ω are to be evaluated at the guiding center position. Finally,

$$\frac{m}{2}v^2 + q\phi(\mathbf{x}) \simeq \frac{m}{2}v_{\parallel}^2 + \mu B + q\phi(\mathbf{X}) \quad (4.12)$$

to first order in \mathbf{r} , where B is to be evaluated at the guiding center position.

From Eqs. (4.9), (4.11), and (4.12), we obtain a new Lagrangian, formulated now with respect to the guiding center coordinates,

$$\mathcal{L} = \left(\frac{q}{c}\mathbf{A} + mv_{\parallel}\mathbf{b} \right) \cdot \dot{\mathbf{X}} + \frac{\mu B}{\Omega}\dot{\phi} - \left(\frac{m}{2}v_{\parallel}^2 + \mu B + q\phi \right), \quad (4.13)$$

where \mathbf{A} and ϕ (as well as B and Ω) depend on the guiding center position \mathbf{X} , not on the particle position \mathbf{x} .

4.2.1.3 Equations of Motion in Guiding Center Coordinates

The Euler-Lagrange equations resulting from Eq. (4.13) read

$$\dot{\mathbf{X}} = v_{\parallel}\mathbf{b} + (B/B_{\parallel}^*) (\mathbf{v}_E + \mathbf{v}_{\nabla B} + \mathbf{v}_c), \quad (4.14)$$

$$\dot{v}_{\parallel} = (q\mathbf{E} - \mu\nabla B) \cdot \dot{\mathbf{X}} / (mv_{\parallel}), \quad (4.15)$$

$$\dot{\mu} = 0, \quad (4.16)$$

$$\dot{\phi} = \Omega, \quad (4.17)$$

where

$$\mathbf{v}_E \equiv \frac{c}{B^2} \mathbf{E} \times \mathbf{B}, \quad (4.18)$$

$$\mathbf{v}_{\nabla B} \equiv \frac{\mu}{m\Omega} \mathbf{b} \times \nabla B , \quad (4.19)$$

$$\mathbf{v}_c \equiv \frac{v_{\parallel}^2}{\Omega} \mathbf{b} \times (\mathbf{b} \cdot \nabla) \mathbf{b} \quad (4.20)$$

are, respectively, the $E \times B$ drift, ∇B drift, and curvature drift and

$$\mathbf{B}^*/B = \mathbf{b} + (v_{\parallel}/\Omega) \nabla \times \mathbf{b} , \quad B_{\parallel}^* = \mathbf{b} \cdot \mathbf{B}^* . \quad (4.21)$$

Some comments are in place here:

- Intuitively, accessible interpretations of the $E \times B$ drift, ∇B drift, and curvature drift are available in many plasma physics textbooks; they shall be omitted here.
- The above equations of motion are correct for (quasi-)static magnetic fields; for time-dependent magnetic fields, one obtains a fourth drift, $\mathbf{v}_t = (v_{\parallel}/\Omega) \mathbf{b} \times \partial_t \mathbf{b}$.
- The prefactor (B/B_{\parallel}^*) is close to unity but must be retained, since it helps to ensure that the structural properties of the equations are preserved.
- The magnetic moment μ is constant in time, reflecting the underlying assumption that the electric and magnetic fields vary slowly in space and time; for an in-depth discussion, see, for example, Ref. [14].
- The gyrophase dynamics is decoupled from the remaining system and can therefore be neglected; this is an important aspect of guiding center theory.

We note that for many space physics applications, the magnetic field can be assumed to be homogeneous; in this case, the equations of motion reduce to

$$\dot{\mathbf{X}} = v_{\parallel} \mathbf{b} + \mathbf{v}_E , \quad \dot{v}_{\parallel} = q E_{\parallel}/m , \quad \dot{\mu} = 0 , \quad \dot{\phi} = \Omega , \quad (4.22)$$

with $E_{\parallel} \equiv \mathbf{b} \cdot \mathbf{E}$. Obviously, this idealization leads to significant simplifications.

Guiding center dynamics has been a long-standing topic in plasma physics, and some readers may be familiar, for example, with the textbook *The Adiabatic Motion of Charged Particles* by T. G. Northrop [15]. Therefore, we'd like to briefly point out some similarities and differences with respect to the approach described above. First, both of these pathways lead to the same results in terms of the final equations of motion. However, the Euler-Lagrange approach is more elegant and compact than the traditional one. Among other things, a complete set of drifts emerges very naturally in the context of the derivation, and the math involved is quite transparent. Second, energy conservation is readily established in the Euler-Lagrange context, while it requires a lot of special attention otherwise. Third, the Euler-Lagrange approach can be extended in a straightforward way to include the dynamics of self-consistent electromagnetic fields, providing a suitable platform for a wide range of modern plasma applications.

Now that we have formulated a model for dynamics of guiding centers in given electromagnetic fields, we will move on to include the effects of fluctuations in these fields. This will take us all the way to gyrokinetics.

4.2.2 Taking into Account Fluctuating Electromagnetic Fields

So far, we have assumed that the electromagnetic fields are prescribed, and we have found that the transition from particle coordinates to guiding center coordinates leads to a useful representation of the particle dynamics. Moreover, it became clear that a Lagrangian approach is a very suitable, transparent, and efficient way to carry out this coordinate change. At this stage, it is natural to ask how this plasma description can be extended to include self-consistent electromagnetic field fluctuations which are created by the plasma particles and which in turn affect their motion. This is the theme of the present subsection.

4.2.2.1 Gyrokinetic Ordering

As was already mentioned in Sect. 4.1.2, gyrokinetic theory is based on a number of assumptions regarding the amplitudes and spatiotemporal scales of the fluctuations in the electromagnetic fields, plasma densities, plasma temperatures, etc. These statements will now be put into a more mathematical form.

Let us assume that each physical quantity of interest can be decomposed into a mean value and a fluctuation, both of which may depend on space and time.² As far as the single-particle distribution function f is concerned, we will denote the mean by f_0 and the fluctuation by f_1 , such that $f = f_0 + f_1$. In practice, f_0 is often taken to be a Maxwellian, but generalizations are possible, and for the present discussion, we will not specify f_0 any further. With this notation, our first assumption can be written as

$$\frac{f_1}{f_0} \sim \epsilon, \quad (4.23)$$

where ϵ is a small parameter. Obviously, this relation suggests that the fluctuations of the plasma densities, plasma temperatures, etc. are also small compared to the respective mean values.

Next, let us turn to the electromagnetic field fluctuations. Assuming that there exists a strong background magnetic field whose magnitude will be denoted by B_0 , we assume that the amplitudes of the magnetic field fluctuations perpendicular and parallel to the mean field, $B_{\perp 1}$ and $B_{\parallel 1}$, are small, i.e.,

$$\frac{B_{\perp 1}}{B_0} \sim \frac{B_{\parallel 1}}{B_0} \sim \epsilon. \quad (4.24)$$

² In the case of the mean values, the spatiotemporal dependencies are assumed to be slow or even absent (in certain situations).

Obviously, the relative amplitudes of $B_{\perp 1}$ and $B_{\parallel 1}$ may depend on various plasma parameters (like the ratio between the plasma pressure and the magnetic pressure, β), and subsidiary orderings are possible; in the present context, we will refrain from discussing such subtleties further, however. The key point is that the magnetic fluctuations obey the same type of smallness relation as the distribution function.

As far as the electric field is concerned, our approach will be slightly different. In particular, we won't take it for granted that a mean electric field exists at all. And as it turns out, a good way to measure the smallness of electric field fluctuations is to compare the associated $E \times B$ drift velocities to the thermal velocity for each particle species. Unless the ions are cold compared to the electrons, this implies that we should focus on the ion thermal velocity. If this constraint is violated, magnetic moment conservation—one of the key foundations of gyrokinetics—is broken [14]. If we allow the electric field fluctuations to also exist at scales comparable to the ion and electron Larmor radii, as stated in Sect. 4.1.2, this implies the relation

$$\frac{q\phi_1}{T_0} \sim \epsilon, \quad (4.25)$$

where T_0 is the mean temperature (of the ions).

Finally, we need to characterize the spatiotemporal scales of all types of fluctuations. In the time domain, all relevant dynamical frequencies ω must be much smaller than the gyrofrequencies Ω of each particle species (in particular, the ions),

$$\frac{\omega}{\Omega} \sim \epsilon. \quad (4.26)$$

Otherwise, gyroaveraging would not make sense in the presence of fluctuations. Typical perpendicular length scales of the fluctuations tend to be small compared to various macroscopic scale lengths L (say, the background density/temperature scale lengths or the scale length of the background magnetic field) but may be comparable to the thermal ion (or electron) Larmor radius ρ , while typical parallel correlation lengths tend to be comparable to L . These properties can be expressed in terms of the typical perpendicular and parallel wave numbers, k_{\perp} and k_{\parallel} . We thus get

$$k_{\perp}\rho \sim k_{\parallel}L \sim 1 \quad (4.27)$$

and hence

$$\frac{k_{\parallel}}{k_{\perp}} \sim \epsilon \quad (4.28)$$

because of $\rho/L \sim \epsilon$, which is necessary to ensure magnetic moment conservation [14]

Again, some comments are in place here:

- Collectively, Eqs. (4.23)–(4.28) define what is often called the gyrokinetic ordering [13]. It provides a basis for various approximations underlying gyrokinetics and helps to define the theory's range of validity.
- For simplicity, we have taken all small parameters to be the same, although they need not be. In fact, many authors introduce separate small parameters and treat them as independent. This provides greater flexibility and transparency.
- Many of the above assumptions can be relaxed, leading to modified gyrokinetic equations—the fundamental ideas always remain the same, however.

In the following, we will discuss how one can derive the gyrokinetic equations of motion on the basis of these assumptions.

4.2.2.2 Gyrokinetic Equations of Motion

Under the conditions described by Eqs. (4.23)–(4.28), the Lagrangian reads

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_1 \quad (4.29)$$

with the zeroth-order term (in ϵ)

$$\mathcal{L}_0 = \left(\frac{q}{c} \mathbf{A}_0 + m v_{\parallel} \mathbf{b}_0 \right) \cdot \dot{\mathbf{X}} + \frac{\mu B_0}{\Omega_0} \dot{\phi} - \left(\frac{m}{2} v_{\parallel}^2 + \mu B_0 + q \phi_0 \right), \quad (4.30)$$

where all spatially dependent quantities are to be evaluated at the guiding center position \mathbf{X} and the first-order term (in ϵ)

$$\mathcal{L}_1 = \frac{q}{c} \mathbf{A}_1(\mathbf{x}, t) \cdot \dot{\mathbf{x}} - q \phi_1(\mathbf{x}, t) \quad (4.31)$$

with $\mathbf{x} = \mathbf{X} + \mathbf{r}$. It should be noted that \mathcal{L}_0 is independent of the gyrophase φ , whereas \mathcal{L}_1 depends on φ via \mathbf{r} . As before, we will remove this effect by averaging \mathcal{L}_1 over φ (from 0 to 2π), a procedure we will denote by $\langle \dots \rangle$. In the following, we will discuss each term in Eq. (4.31) separately.

First, we find that

$$\langle \mathbf{A}_1(\mathbf{X} + \mathbf{r}, t) \cdot \dot{\mathbf{X}} \rangle = \langle A_{\parallel 1}(\mathbf{X} + \mathbf{r}) \rangle \mathbf{b}_0 \cdot \dot{\mathbf{X}} \quad (4.32)$$

with $A_{\parallel 1} = \mathbf{b} \cdot \mathbf{A}_1$. Here, the term $\langle A_{\parallel 1}(\mathbf{X} + \mathbf{r}) \rangle$ means that $A_{\parallel 1}$ is *not* to be evaluated at the guiding center position \mathbf{X} ; instead, it is to be averaged along the trajectory of a quickly gyrating particle whose guiding center position is \mathbf{X} . If one expands $A_{\parallel 1}$ into an infinite Taylor series about \mathbf{X} and gyroaverages this expression, one obtains

$$\langle A_{\parallel 1}(\mathbf{X} + \mathbf{r}, t) \rangle = J_0(\lambda) A_{\parallel 1}(\mathbf{X}, t) \equiv \tilde{A}_{\parallel 1}(\mathbf{X}, \mu, t), \quad (4.33)$$

where J_0 is the zeroth-order Bessel function of the first kind (which only depends on even powers of λ) and $\lambda^2 = -r^2\nabla_\perp^2$ with the perpendicular Laplacian ∇_\perp^2 . It should be noted that the gyroaveraging procedure introduces a dependence on the magnetic moment μ , which sets the Larmor radius.

In an analogous fashion, one obtains

$$\frac{q}{c}\langle \mathbf{A}_1(\mathbf{X} + \mathbf{r}, t) \cdot \dot{\mathbf{r}} \rangle = -(2/\lambda) J_1(\lambda) \mu B_{\parallel 1}(\mathbf{X}, t) \equiv -\mu \bar{B}_{\parallel 1}(\mathbf{X}, \mu, t), \quad (4.34)$$

where J_1 is the first-order Bessel function of the first kind (which only depends on odd powers of λ), and

$$\langle \phi_1(\mathbf{X} + \mathbf{r}, t) \rangle = J_0(\lambda) \phi_1(\mathbf{X}, t) \equiv \bar{\phi}_1(\mathbf{X}, \mu, t). \quad (4.35)$$

Consequently, upon gyroaveraging, \mathcal{L}_1 as given by Eq. (4.31) becomes

$$\mathcal{L}_1 = \left(\frac{q}{c} \bar{A}_{\parallel 1} \mathbf{b}_0 \right) \cdot \dot{\mathbf{X}} - (\mu \bar{B}_{\parallel 1} + q \bar{\phi}_1), \quad (4.36)$$

where we have suppressed the variable dependencies to simplify the notation. Together, Eqs. (4.30) and (4.36) define the relevant Lagrangian and determine the gyrokinetic equations of motion in fluctuating electromagnetic fields.

The corresponding Euler-Lagrange equations for the five gyrokinetic variables $(\mathbf{X}, v_{\parallel}, \mu)$ read

$$\dot{\mathbf{X}} = v_{\parallel} \mathbf{b} + (B_0/B_{\parallel}^*) (\mathbf{v}_{m1} + \mathbf{v}_{E0} + \mathbf{v}_{E1} + \mathbf{v}_{\nabla B0} + \mathbf{v}_{\nabla B1} + \mathbf{v}_{c0}), \quad (4.37)$$

$$\dot{v}_{\parallel} = (q \mathbf{E}_0 + q \bar{\mathbf{E}}_1 - \mu \nabla B_0 - \mu \nabla \bar{B}_{\parallel 1}) \cdot \dot{\mathbf{X}} / (mv_{\parallel}), \quad (4.38)$$

$$\dot{\mu} = 0, \quad (4.39)$$

where \mathbf{v}_{E0} , $\mathbf{v}_{\nabla B0}$, and \mathbf{v}_{c0} are given, respectively, by Eqs. (4.18), (4.19), and (4.20), and

$$\mathbf{v}_{m1} = v_{\parallel} (\bar{\mathbf{B}}_{\perp 1}/B_0), \quad (4.40)$$

$$\mathbf{v}_{E1} = \frac{c}{B^2} \bar{\mathbf{E}}_1 \times \mathbf{B}, \quad (4.41)$$

$$\mathbf{v}_{\nabla B1} = \frac{\mu}{m\Omega} \mathbf{b} \times \nabla \bar{B}_{\parallel 1} \quad (4.42)$$

with $\bar{\mathbf{E}}_1 = -\nabla\bar{\phi}_1 - \mathbf{b}\partial_t\bar{A}_{\parallel 1}/c$, $\bar{\mathbf{B}}_{\perp 1} = \nabla \times (\bar{A}_{\parallel 1}\mathbf{b})$, and $\mathbf{b} \equiv \mathbf{B}_0/B_0$. The term \mathbf{v}_{m1} describes the motion perpendicular to the background magnetic field due to particles following fluctuating field lines, while the terms \mathbf{v}_{E1} and $\mathbf{v}_{\nabla B1}$ describe, respectively, contributions to the $E \times B$ drift and ∇B drift due to fluctuations in the electrostatic potential and parallel magnetic field.

Let us again provide some comments:

- It is interesting to note that as a consequence of the fast gyromotion perpendicular to the background magnetic field, the magnetic field fluctuations are completely characterized by two (not three!) scalar quantities, namely, $A_{\parallel 1}$ and $B_{\parallel 1}$. From the former, the two perpendicular components of the fluctuating magnetic field are readily derived.
- The perpendicular drifts associated with the electromagnetic field fluctuations can be interpreted in the usual way. The main novelty is that to evaluate the respective potentials, one needs to take into account that particles “feel” them along their gyroorbit, not at the guiding center position.
- The magnetic moment is again conserved, even in the presence of electromagnetic field fluctuations.

In the important special case of a homogeneous background magnetic field, the gyrokinetic equations of motion become

$$\dot{\mathbf{X}} = v_{\parallel}\mathbf{b} + \mathbf{v}_{m1} + \mathbf{v}_{E0} + \mathbf{v}_{E1} + \mathbf{v}_{\nabla B1}, \quad (4.43)$$

$$\dot{v}_{\parallel} = (q\mathbf{E}_0 + q\bar{\mathbf{E}}_1 - \mu\nabla\bar{B}_{\parallel 1}) \cdot \dot{\mathbf{X}}/(mv_{\parallel}), \quad (4.44)$$

$$\dot{\mu} = 0. \quad (4.45)$$

We note that there remain contributions associated with $\nabla\bar{B}_{\parallel 1}$, despite $\nabla B_0 = 0$.

4.2.2.3 Gyrokinetic Vlasov-Maxwell Equations

At this stage, we would like to point out that the following paragraphs involve some concepts from more advanced classical mechanics, which transcend the material typically included in introductory courses. The reader may therefore not be familiar with them. In this context, we highly recommend the review article by A. J. Brizard and T. S. Hahm [7]. Therein, the mathematical foundations of gyrokinetics are clearly laid out, and a large number of references are provided. If one is interested in fully appreciating and mastering gyrokinetics, this review article is an excellent starting point.

Having said that, it is time to move on to deriving the complete set of self-consistent gyrokinetic Vlasov-Maxwell equations, based on the equations of

motions just discussed. In order to be able to do that, however, an important point needs to be addressed first, namely, the notion of gyrocenter coordinates.

From Guiding Center Coordinates to Gyrocenter Coordinates

When we derived the description of charged particle dynamics in prescribed (quasi-)static electromagnetic fields in Sect. 4.2.1, we introduced the guiding center coordinates $\mathbf{Z} \equiv (\mathbf{X}, v_{\parallel}, \mu, \varphi)$. The corresponding Lagrangian, \mathcal{L}_0 , is independent of the gyrophase φ . Upon introducing electromagnetic field fluctuations, a gyrophase dependence is reintroduced via the fact that ϕ_1 and \mathbf{A}_1 in \mathcal{L}_1 are to be evaluated at the particle position $\mathbf{x} = \mathbf{X} + \mathbf{r}$, with \mathbf{r} being a function of φ . In Sect. 4.2.2.2, we removed this φ dependence by simply averaging \mathcal{L}_1 over one gyroperiod.

As it turns out, a more appropriate way to achieve the very same goal is to carry out a near-identity coordinate transform from guiding center to the so-called *gyrocenter coordinates* $\bar{\mathbf{Z}}$. The mathematical relationship between these two sets of coordinates is given by

$$\bar{\mathbf{Z}} = \mathbf{Z} + \{\tilde{S}_1(\mathbf{Z}, t), \mathbf{Z}\} \quad (4.46)$$

to first order in the small parameter ϵ , where $\{\cdot, \cdot\}$ denotes the Poisson bracket. Here, the so-called *first-order generating function* $\tilde{S}_1(\mathbf{Z}, t)$ is determined by

$$\frac{\partial \tilde{S}_1}{\partial t} + \{\tilde{S}_1, H_0\} = H_1 - \langle H_1 \rangle \equiv \tilde{H}_1 \quad (4.47)$$

with the zeroth-order Hamiltonian

$$H_0 = \frac{m}{2}v_{\parallel}^2 + \mu B_0 + q\phi_0 \quad (4.48)$$

and the first-order Hamiltonian

$$H_1 = -\mathcal{L}_1 = q\phi_1(\mathbf{x}, t) - \frac{q}{c}\mathbf{A}_1(\mathbf{x}, t) \cdot \dot{\mathbf{x}} \quad (4.49)$$

with $\mathbf{x} = \mathbf{X} + \mathbf{r}$. To leading order in ϵ , Eq. (4.47) reduces to $\Omega_0 \partial_{\varphi} \tilde{S}_1 = \tilde{H}_1$, which has the solution

$$\tilde{S}_1(\mathbf{Z}, t) = \frac{1}{\Omega_0(\mathbf{X}, t)} \int d\varphi \tilde{H}_1(\mathbf{Z}, t), \quad (4.50)$$

subject to the condition $\langle \tilde{S}_1 \rangle = 0$. According to these relationships, $\bar{\mathbf{Z}} = \mathbf{Z} + O(\epsilon^2)$, which in hindsight justifies the use of the \mathbf{Z} coordinates in the derivation of \mathcal{L}_1 in Sect. 4.2.2.2.

Gyrokinetic Vlasov Equation

An important consequence of the transition from particle coordinates \mathbf{z} to guiding center coordinates \mathbf{Z} to gyrocenter coordinates $\bar{\mathbf{Z}}$ is that the functional form of the single-particle distribution function changes such that

$$f(\mathbf{z}, t) = f_g(\mathbf{Z}, t) = f_G(\bar{\mathbf{Z}}, t). \quad (4.51)$$

Moreover,

$$d^6\mathbf{z} = J_g d^6\mathbf{Z} = J_G d^6\bar{\mathbf{Z}} \quad (4.52)$$

with the Jacobian $J_g = J_G = B_{\parallel}^*/m$. The gyrokinetic Vlasov equation in gyrocenter coordinates can thus be written in conservation law form,

$$\frac{\partial}{\partial t} (J_G(\bar{\mathbf{Z}}, t) f_G(\bar{\mathbf{Z}}, t)) + \frac{\partial}{\partial \bar{\mathbf{Z}}} \cdot (\dot{\bar{\mathbf{Z}}} J_G(\bar{\mathbf{Z}}, t) f_G(\bar{\mathbf{Z}}, t)) = 0, \quad (4.53)$$

or in advection form,

$$\left(\frac{\partial}{\partial t} + \dot{\bar{\mathbf{Z}}} \cdot \frac{\partial}{\partial \bar{\mathbf{Z}}} \right) f_G(\bar{\mathbf{Z}}, t) = 0, \quad (4.54)$$

thanks to the fact that the phase-space flow is incompressible,

$$\frac{\partial}{\partial \bar{\mathbf{Z}}} \cdot (\dot{\bar{\mathbf{Z}}} J_G(\bar{\mathbf{Z}}, t)) = 0. \quad (4.55)$$

Here, $\dot{\bar{\mathbf{Z}}}$ can be replaced by $\dot{\mathbf{Z}}$ as given in Sect. 4.2.2.2 to first order in ϵ , as mentioned above.

Gyrokinetic Maxwell's equations

To complement the gyrokinetic Vlasov equation, we now need to derive a gyrokinetic version of Maxwell's equations to determine the self-consistent electromagnetic fields. Let us first focus on the particle density n for any given species, which is needed to compute the charge density and thus the electric field. In \mathbf{z} coordinates, we have $n(\mathbf{x}, t) = \int d^3\mathbf{v} f(\mathbf{x}, \mathbf{v}, t)$, and in $\bar{\mathbf{Z}}$ coordinates, we get

$$n(\mathbf{x}, t) = \int d^6\bar{\mathbf{Z}} \delta(\bar{\mathbf{X}} + \mathbf{r} - \mathbf{x}) J_G(\bar{\mathbf{Z}}, t) f_G(\bar{\mathbf{Z}}, t). \quad (4.56)$$

Here, to first order in ϵ , we have

$$f_G(\bar{\mathbf{Z}}, t) = f_g(\bar{\mathbf{Z}}, t) + \{\tilde{S}_1(\bar{\mathbf{Z}}, t), f_g(\bar{\mathbf{Z}}, t)\}, \quad (4.57)$$

where it can be shown that the second term on the right-hand side captures polarization and magnetization effects in the plasma. Linearizing this expression by replacing f_g in the Poisson bracket with a suitable f_0 , we obtain

$$f_G = f_g + \frac{\Omega_0}{B_0} \tilde{A}_{\parallel 1} \frac{\partial f_0}{\partial v_{\parallel}} + \left(q\tilde{\phi}_1 - \frac{q}{c} v_{\parallel} \tilde{A}_{\parallel 1} - \mu \bar{B}_{\parallel 1} \right) \frac{1}{B_0} \frac{\partial f_0}{\partial \mu}. \quad (4.58)$$

Taking f_0 to be an unshifted Maxwellian, we arrive at

$$n_1(\mathbf{x}, t) = \bar{n}_1(\mathbf{x}, t) - \int d^6 \bar{\mathbf{Z}} \delta(\bar{\mathbf{X}} + \mathbf{r} - \mathbf{x}) J_G \left(q\tilde{\phi}_1 - \mu \bar{B}_{\parallel 1} \right) \frac{f_0}{T_0}, \quad (4.59)$$

where the first term is

$$\bar{n}_1(\mathbf{x}, t) \equiv \int d^6 \bar{\mathbf{Z}} \delta(\bar{\mathbf{X}} + \mathbf{r} - \mathbf{x}) J_G(\bar{\mathbf{Z}}, t) f_{g1}(\bar{\mathbf{Z}}, t) \quad (4.60)$$

and can be interpreted as the particle density one gets by adding the contributions from all particle trajectories (with gyrocenter position $\bar{\mathbf{X}}$) passing through \mathbf{x} . Moreover, $\tilde{\phi}_1(\bar{\mathbf{Z}}, t) \equiv \phi_1(\mathbf{x}, t) - \bar{\phi}_1(\bar{\mathbf{Z}}, t)$ with $\mathbf{x} = \bar{\mathbf{X}} + \mathbf{r}$ and $\mathbf{r} = \mathbf{r}(\bar{\mathbf{Z}})$. $q\phi_1/T_0$ can be pulled out of the integral, leaving behind the zeroth moment of f_0 , which we denote by n_0 . After some cumbersome but straightforward algebra, one arrives at the expression (in \mathbf{x} coordinates)

$$\frac{n_1}{n_0} = \frac{\bar{n}_1}{n_0} - (1 - \Gamma_0) \frac{q\phi_1}{T_0} + (\Gamma_0 - \Gamma_1) \frac{B_{\parallel 1}}{B_0} \quad (4.61)$$

in terms of $\Gamma_n(b) = I_n(b)\exp(-b)$ with the n th-order modified Bessel functions I_n and $b \equiv -r^2\nabla_{\perp}^2$. This expression can be used in Poisson's equation,

$$\nabla_{\perp}^2 \phi_1 = -4\pi \sum_s q_s n_{1s} \equiv -4\pi \rho_c, \quad (4.62)$$

to compute the electrostatic potential. Here, we have assumed $k_{\parallel} \ll k_{\perp}$ and $\sum_s q_s n_{0s} = 0$. In an analogous fashion, one can derive the relationships

$$\nabla_{\perp}^2 A_{\parallel 1} = -\frac{4\pi}{c} \sum_s q_s \Gamma_{\parallel 1s} \equiv -\frac{4\pi}{c} J_{\parallel 1} \quad (4.63)$$

and

$$\frac{B_{\parallel 1}}{B_0} = - \sum_s \beta_s \left[\frac{\bar{p}_{\perp 1s}}{p_{0s}} + (\Gamma_0 - \Gamma_1) \frac{q\phi_1}{T_{0s}} + 2b_s (\Gamma_0 - \Gamma_1) \frac{B_{\parallel 1}}{B_0} \right] \quad (4.64)$$

with $p_{0s} = n_{0s} T_{0s}$ in terms of the parallel particle flux

$$\Gamma_{\parallel 1}(\mathbf{x}, t) \equiv \int d^6 \bar{\mathbf{Z}} \delta(\bar{\mathbf{X}} + \mathbf{r} - \mathbf{x}) J_G(\bar{\mathbf{Z}}, t) v_{\parallel} f_{g1}(\bar{\mathbf{Z}}, t) \quad (4.65)$$

and the perpendicular pressure

$$\bar{p}_{\perp 1}(\mathbf{x}, t) \equiv \int d^6 \bar{\mathbf{Z}} \delta(\bar{\mathbf{X}} + \mathbf{r} - \mathbf{x}) J_G(\bar{\mathbf{Z}}, t) \mu B_0(\bar{\mathbf{Z}}, t) (2J_1(b)/b) f_{g1}(\bar{\mathbf{Z}}, t). \quad (4.66)$$

This completes the derivation of the gyrokinetic equations.

4.2.3 Further Reading

The goal of the present “primer on gyrokinetics” is to serve as an introduction to the topic, also for the novice. Therefore, a deliberate attempt has been made to provide a relatively direct, “light-weight” approach in terms of the mathematical techniques involved. A solid background in classical mechanics and basic plasma physics should suffice to appreciate and grasp the material laid out above. (If necessary, the reader can consult a large number of excellent textbooks in these two areas.)

Building on earlier work on *linear* gyrokinetics, a large number of publications on *nonlinear* gyrokinetics have emerged since the seminal paper by Frieman and Chen in 1982 [13]. In this context, gyrokinetic theory has been put on a firm and transparent mathematical foundation, insightful interpretations of its physical content have been provided, and the original theory has gradually been extended and generalized to make it applicable to a growing number of fusion, laboratory, astrophysical, and space plasmas. In the following, a few particularly important publications on these issues will be recommended for further reading.

Arguably, the best starting point for a deeper engagement with gyrokinetic theory is the review article by Brizard and Hahm from 2007 [7]. It gives a rather detailed account of the underlying mathematical formalisms and provides numerous pointers to the original literature that modern gyrokinetic theory is built on. A second useful reference is the review paper by Cary and Brizard from 2009 [8]. It nicely lays out the many advantages of a Hamiltonian (or Lagrangian) approach for developing reduced dynamical equations for charged particles in slowly varying electromagnetic fields. A third review to mention here is the one by Krommes from 2012 [9]. Among other things, it puts gyrokinetic theory into some historical and conceptual context, and maybe even more importantly, it discusses some basic physical properties of gyrokinetic systems, in particular in relation to microturbulence in magnetized plasmas.

These three review articles can serve as a starting point to familiarize oneself with some of the original literature which also deals with advanced topics in gyrokinetics including but not limited to

- Gyrokinetic field theory
- Gyrokinetic energy and momentum conservation laws
- Gyrokinetic collision operators
- High-frequency extensions of gyrokinetics
- Gyrokinetic waves and instabilities
- Gyrokinetic turbulence
- Gyrokinetic magnetic reconnection

Although gyrokinetics has matured significantly over the years, it remains a dynamic and vibrant area of research to this day. Moreover, it offers excellent perspectives for innovative and fascinating future work, also and especially in space plasma physics. A look into the recent and current literature illustrates this point.

4.3 Computational Gyrokinetics

For simple magnetic geometries (like a homogeneous magnetic field), simple simulation domains (like a triply periodic cuboid), and simple problems (like linearized equations), analytic or semi-analytic treatments are sometimes feasible. In more complex situations, however, one generally has to resort to numerical solutions of the gyrokinetic equations. Here, the goal is usually twofold: to better understand the very nature of certain physical processes on a qualitative level and to interpret or predict certain experiments/observations in a quantitative way. In the following, the two leading techniques in computational gyrokinetics are briefly discussed.

4.3.1 The Lagrangian (Particle-in-Cell) Approach

The particle-in-cell (PIC) approach is very popular in the context of fully kinetic simulations of weakly collisional plasmas on the basis of the Vlasov-Maxwell (or similar) equations (see Chap. 6 of this book). In fact, so much so that the words “PIC” and “kinetic” are sometimes used interchangeably in respective publications or presentations. It should be kept in mind, however, that, strictly speaking, “kinetic” refers to the underlying physical model, while “PIC” refers to a numerical technique. In particular since we are going to describe alternative computational methods below, we suggest a more careful use of the “PIC” label.

In the context of nonlinear gyrokinetics, the PIC approach was pioneered by Lee and co-workers [16] shortly after the equations were first written down. Based on a Lagrangian interpretation of the gyrokinetic Vlasov equation, it exploits the fact that in the collisionless limit, the value of the gyrocenter distribution function of

any given particle species does not change in time along phase-space characteristics. Thus, the idea is to sample the distribution function at an initial time by a sufficiently large number of the so-called markers (which represent a discrete set of gyrocenters) and then track their trajectories. As usual in the context of a PIC approach, at each time step, the charge and current densities are gathered onto a spatial grid, the field equations are solved on that same grid, and the resulting updated values of the fields are used to further advance the marker positions.

The PIC approach can be interpreted as a Monte Carlo sampling method [17]. This implies that it is subject to numerical noise which can dominate PIC simulations and lead to wrong results [18]. For this reason, noise reduction/control is a key issue in gyrokinetic PIC simulation, and much work has been dedicated to address it. The key challenge is that the relative noise level decreases with increasing particle number N only as $N^{-1/2}$; this puts severe limits on convergence tests in many practical applications. However, it has been demonstrated that the prefactor in this scaling can be reduced, for example, via a so-called δf method [19, 20]. Here, the gyrocenter distribution function f is split into a (Maxwellian) background and a small fluctuation δf (which we denoted by f_1 above), and only δf is represented by the markers.

In summary, the PIC approach has several appealing aspects, including its intuitive nature, its relative ease of implementation, and its suitability for high-performance computing. On the other hand, the noise problem needs to be carefully monitored and controlled in each simulation, if necessary via a substantial increase of the particle number N [21]. Several gyrokinetic PIC codes exist and are widely used in fusion research and beyond [22–27].

4.3.2 The Eulerian (Grid-Based) Approach

Throughout much of the 1980s and 1990s, computational gyrokinetics relied exclusively on the PIC approach. Toward the late 1990s, however, various efforts were started to explore the applicability of a grid-based approach. Here, one adopts a Eulerian interpretation of the gyrokinetic Vlasov equation and often employs numerical methods originally developed in the computational fluid dynamics (CFD) and related communities. One discretizes the gyrocenter distribution function(s) on a fixed grid in a five-dimensional phase space and uses (a mix of) spectral, finite-difference, finite-volume, and/or finite-element methods to represent the gyrokinetic equations on this grid. The time stepping is often realized via time-explicit higher-order (e.g., Runge-Kutta) schemes, based on the method of lines.

Originally thought to be computationally prohibitive, the grid-based approach to gyrokinetics has turned out to be at least as efficient as the PIC approach, provided the algorithms are carefully selected and implemented. Moreover, the noise problem is absent, thus removing a major source of concern. Not surprisingly then, since their inception in 2000, grid-based nonlinear gyrokinetic codes like GENE [28] and GS2 [29] gradually took a lead in exploring various aspects of microturbulence and magnetic reconnection in magnetized fusion, laboratory, astrophysical, and

space plasmas. In the case of GENE, this trend was helped significantly by the code's public availability (<http://genecode.org>), its state-of-the-art implementation on many of the world's largest supercomputers, and a team-based effort to provide high-quality user support. Some examples of successful applications of GENE to space physics problems will be given in Sect. 4.4.

Another, more recently developed grid-based approach to computational gyrokinetics is the semi-Lagrangian method. Here, in the collisionless limit, the values of the gyrocenter distribution functions on phase-space grid points at the next time step are computed by tracing back the respective characteristics to their positions at the present time step and using interpolations to obtain the required values. While this method has been used successfully in the context of fusion research, it has yet to be applied to problems in space plasma physics [30].

In summary, the grid-based approach tends to be somewhat harder to implement than a PIC approach but has been shown to lead to simulation codes that are very competitive in terms of time-to-solution and parallel scalability. In addition, grid-based codes don't suffer from a noise problem, thus removing a major source of concern regarding PIC codes. Eulerian gyrokinetic codes have been playing a key role in fusion research (and beyond) over the last two decades or so [28, 29, 31–34].

4.4 Applications of Gyrokinetics to Solar Wind Turbulence

As mentioned in Sect. 4.4, over the last decade or so, gyrokinetics has found its way into many space plasma physics applications. In the following, we will focus on a particularly timely and interesting example, namely, the physics of the kinetic range of solar wind turbulence. Two key questions to be addressed in this context are:

- What is the character of solar wind turbulence at kinetic scales?
- How can turbulent energy be dissipated in the almost collisionless solar wind?

4.4.1 On the Nature of Solar Wind Turbulence at Kinetic Scales

The temperature of the solar wind as a function of the distance from the Sun, as measured by spacecraft, can only be explained by the presence of turbulent heating throughout the heliosphere. This is one of the reasons why a deeper understanding of the physical mechanisms that dissipate small-scale turbulent fluctuations, ultimately converting the turbulent energy into heat, is considered one of the major unsolved problems in the space plasma physics and plasma astrophysics communities [35, 36].

4.4.1.1 Different Kinetic Descriptions: Advantages and Drawbacks

The solar wind constitutes an excellent “laboratory” to study the nature of space plasma turbulence *in situ* via spacecraft observations. In this context, it is generally assumed that at long wavelengths, the turbulent dynamics can be well described by the theory of magnetohydrodynamics (MHD). With more and more observational

data of solar wind turbulence at kinetic scales becoming available over the last several years, systematic attempts to use kinetic simulation to reproduce some key features of the small-scale turbulent fluctuations became an enormous research opportunity, attracting a lot of attention. The published work tends to fall into three categories, based on the underlying physical model:

- Fully kinetic simulations
- Hybrid kinetic-fluid simulations
- Gyrokinetic simulations

Fully kinetic simulations, as described in Chaps. 5 and 6 of this book, provide the most complete description of the system under consideration, of course, but well-resolved six-dimensional simulations using realistic physical parameters are still computationally prohibitive as of today. In practice, to make fully kinetic studies feasible, one usually reduces the numerical resolution and/or simulation domain size, and one adapts the physical parameters (e.g., one reduces the proton-to-electron mass ratio significantly, thus compressing the dynamical scale range). Therefore, the advantage of employing a high-fidelity physical model is diminished by the need to work with unrealistic physical settings and/or moderately sized simulation domains.

Hybrid kinetic-fluid simulations provide an attractive alternative to fully kinetic studies. Given that they are based on a simple fluid model for the electron dynamics, the computational cost is reduced substantially, making such simulations much more feasible. On the other hand, as already pointed out in Sect. 4.3, important kinetic electron physics like electron Landau damping is missing from this reduced plasma description. This is a serious drawback if one wants to investigate the different pathways toward turbulent dissipation. Attempts to incorporate electron Landau damping in an *ad hoc* fashion have to balance fidelity with efficiency.

Gyrokinetic simulations turn out to provide a third possible approach to study solar wind turbulence at kinetic scales. According to the theory of MHD turbulence, one expects that the turbulent fluctuations which are driven near the Sun in a more or less isotropic fashion become more and more anisotropic as they undergo a turbulent cascade toward smaller scales. Here, the fluctuations tend to be stretched out along the mean magnetic field created by the large-scale motion of the solar wind according to the notion of critical balance between the linear wave frequencies and nonlinear decorrelation rates which suggests $k_{\parallel} \propto k_{\perp}^{2/3}$, thus establishing $k_{\parallel} \ll k_{\perp}$. At the same time, the fluctuation amplitudes gradually drop to small values, such that $f_1 \ll f_0$. Moreover, the observed fluctuation frequencies usually tend to be small compared to the ion cyclotron frequency, i.e., $\omega \ll \Omega$. Taken together, these statements imply that the tail of the turbulent cascade in the solar wind appears to be in line with the gyrokinetic ordering, suggesting that gyrokinetic simulation should be able to shed some light on the above questions.

Needless to say that despite their respective limitations, all three of these approaches are pursued in parallel to interpret and better understand the available

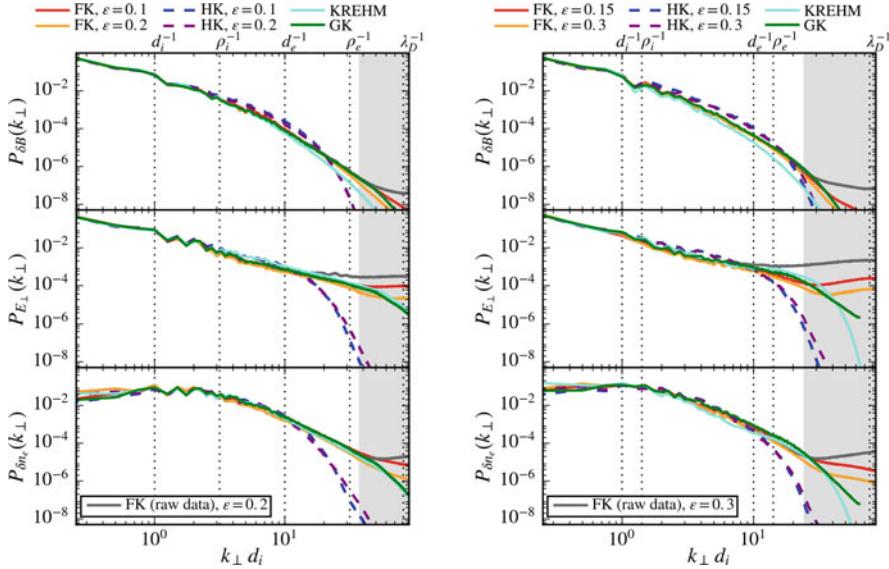


Fig. 4.3 One-dimensional turbulent energy spectra for $\beta_i = 0.1$ (left) and $\beta_i = 0.5$ (right) based on fully kinetic (“FK”), hybrid kinetic (“HK”), and gyrokinetic (“GK”) simulations. Figures from [12] reproduced with permission by AAS. Shown from top to bottom are the spectra of the magnetic field, electric field, and electron density. The gray-shaded regions indicate the range where the PIC noise dominates over the turbulent fluctuations for the short-time-averaged FK PIC data

observational data from the solar wind. Systematic comparisons between them have also been carried out, as exemplified in Fig. 4.3. In the following, we will focus our attention on gyrokinetic studies.

4.4.1.2 Gyrokinetic Simulations of Solar Wind Turbulence

As discussed above, gyrokinetics is well suited to study the kinetic scales of solar wind turbulence which act as the dissipation range of the MHD cascade [10]. Here, the energy in the turbulent fluctuations arriving from larger scales is used to heat the ions and electrons of the plasma. At the same time, other plasma processes take place, including magnetic reconnection and particle acceleration (i.e., the creation of suprathermal particles). All of these phenomena are inherently kinetic in nature.

A useful starting point for a comparison between observations and gyrokinetic simulations is the perpendicular (w.r.t. the mean magnetic field) wave number spectra of various fluctuating quantities. A first study along these lines was presented in 2008 by Howes and co-workers [37]. Using a homogeneous background magnetic field and physical parameters representative of the solar wind at about 1 AU ($\beta_i = 1$ and $T_i/T_e = 1$), the transition from MHD scales to kinetic scales at $k_{\perp} \rho_i \sim 1$ was investigated. In the MHD limit, $k_{\perp} \rho_i \ll 1$, gyrokinetics turns into reduced MHD, and consequently one expects gyrokinetic simulations to be consistent with MHD

simulations in this regime. This was confirmed in Ref. [37], where the electric and perpendicular magnetic energy spectra both showed a scaling consistent with $k^{-5/3}$ in a scale range of $k_{\perp}\rho_i \sim 0.1 - 1$. In another simulation, focused on the scale range of $k_{\perp}\rho_i \sim 0.5 - 10$, these energy spectra exhibited a break around the ion gyroscale and spectral slopes in line with scaling predictions for critically balanced turbulence of kinetic Alfvén waves below the ion gyroscale (scalings of about $k_{\perp}^{-1/3}$ and $k_{\perp}^{-7/3}$ for electric and perpendicular/parallel magnetic fluctuations, respectively). This behavior is roughly consistent with in situ turbulence measurements. This finding was confirmed in a follow-up simulation using the real proton-to-electron mass ratio and covering a scale range of $k_{\perp}\rho_i \sim 1 - 40$ [38].

A much larger gyrokinetic simulation of solar wind turbulence was carried out in 2015 by Told and co-workers by means of the GENE code [39]. Here, for the very first time, the entire scale range from MHD scales down to the electron gyroradius, $k_{\perp}\rho_i \sim 0.2 - 50$, was covered in a single simulation (using the real proton-to-electron mass ratio). This implies, in particular, that the physics at the spectral break around $k_{\perp}\rho_i \sim 1$ was fully included in the simulation and well separated from the drive range at $k_{\perp}\rho_i \sim 0.2$. In the range $k_{\perp}\rho_i < 1$, MHD-type spectra were observed, and for $k_{\perp}\rho_i > 1$, all quantities exhibited rather well-defined power law spectra, consistent with scaling predictions for critically balanced turbulence of kinetic Alfvén waves and with the previous work as described above (see Fig. 4.4). These findings lend confidence to the validity of the results regarding the spectral behavior.

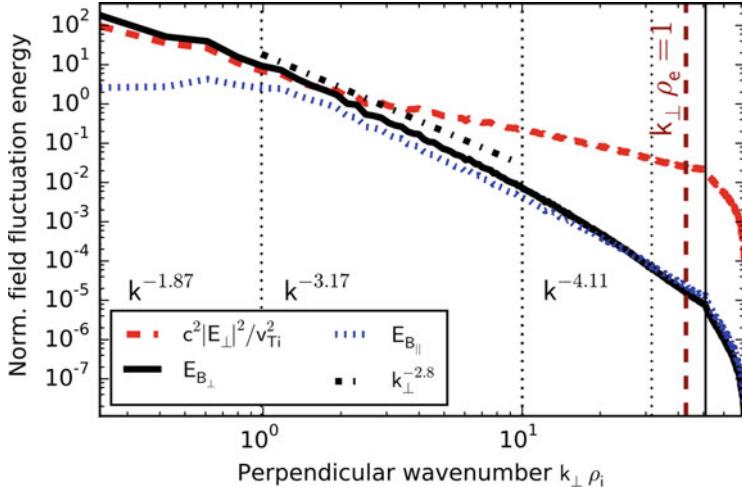


Fig. 4.4 Normalized energy spectra from a large-scale gyrokinetic turbulence simulation down to the electron gyroradius scale. Reprinted figure with permission from [39]; copyright (2015) by the American Physical Society. Power law exponents obtained from the B_{\perp} energy spectra within the dotted sections are printed into the plot

Further analysis focused on the energy transfer at kinetic scales. In this context, it was demonstrated that there exists a local direct cascade in perpendicular wave number space which is mediated to a significant degree by modes in the tail of the MHD range (thus highlighting the fact that kinetic range tends to act as an “entity”) [39]. Another key finding was that for these physical parameters, about 70% of the injected energy is dissipated by the electrons and only about 30% by the ions. Here, the electron dissipation spectrum peaks around $k_{\perp}\rho_i \sim 1 - 5$, while the ion counterpart peaks at much smaller scales, namely, around $k_{\perp}\rho_i \sim 25$ [39]. These results highlight the presence of strong dissipation throughout the entire kinetic range, justifying the common use of the expression “dissipation range.”

The GENE-based study just discussed has been further extended in Ref. [40], with an emphasis on understanding the heating processes in more details. Previous (fully) kinetic plasma simulations indicated that the turbulent dissipation in space plasmas is highly inhomogeneous and occurs predominantly in current sheets. However, this conjecture was not based on calculations of the actual local heating rate. Instead, it depended on proxies like the so-called electron-frame dissipation measure (EFDM) $D_e = \mathbf{J} \cdot (\mathbf{E} + \mathbf{v}_e \times \mathbf{B})$, since no explicit expression for collisional dissipation was available. The gyrokinetic simulations with GENE, on the other hand, included a linearized Landau-Boltzmann collision operator, allowing for a much more detailed and reliable analysis of the dissipation processes. It is important to note that any kind of irreversible dynamics associated with plasma heating must be due to collisions. Kinetic processes like Landau damping merely play the role of a catalyst, enhancing the creation of small-scale structures in velocity space which are susceptible to dissipative effects even in weakly collisional regimes. To assess this kind of physics, collisions must be included in the simulations (Fig. 4.5).

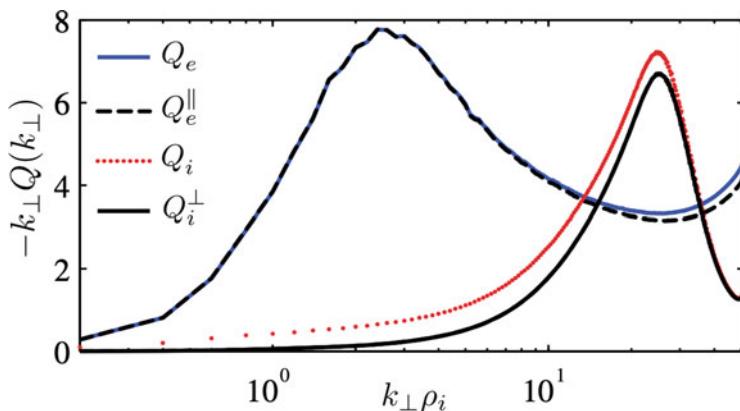


Fig. 4.5 Normalized perpendicular wave number spectra for electron and ion heating rates for the simulation shown in Fig. 4.4. Reprinted figure with permission from [40]; copyright (2016) by the American Physical Society. The parallel electron and perpendicular ion contributions to the heating rate are shown for comparison. For the electrons, *parallel* phase mixing (i.e., Landau damping) dominates; for the ions, it is *perpendicular* phase mixing.

The following conclusions could be drawn from this study:

- The real space structures of ion and electron heating differ fundamentally. While the former is more uniform, the latter exhibits a rather patchy nature.
- While an EFDM analysis of the simulation data suggests that for the present parameters about 90% of the electron/ion heating takes place in current sheets, measurements of the local heating rates reveal that the real values are only about 40%. In other words, dissipation happens mostly outside current sheets, and the EFDM results are misleading.
- The energy transfer from fields to particles (as measured by D_e) is necessary for dissipation to occur; however, under weakly collisional conditions, the conversion into heat need not happen at the same location in space and time as this transfer, due to the phase-space cascade.
- While *linear, parallel* phase mixing (i.e., electron Landau damping) is the preferred route to dissipation for electron heating, *nonlinear, perpendicular* phase mixing (via μ -dependent $\mathbf{E} \times \mathbf{B}$ motion) is the main catalyst for ion heating. These processes create small-scale structures in v_{\parallel} and μ space, respectively.

In follow-up studies, several fundamental aspects of gyrokinetic turbulence for solar-wind-type parameters have been investigated in more details. First, a systematic analysis of nonlinear energy transfers was performed, based on the notion of the so-called locality functions [41]. It could be shown that despite the fact that the nonlinear interactions in gyrokinetic turbulence are nonlocal in nature, they possess an asymptotic local component. This implies that such systems display certain aspects of universality, as observed in hydrodynamic turbulence, although their physics is more involved overall. Second, the nature of phase-space intermittency was studied, finding that the electron dynamics is strongly intermittent compared to weakly intermittent protons [42]. In this context, links between intermittency and electron Landau damping were established, as intermittent electron structures also exhibit strong parallel velocity structures. Third, a new method was proposed to detect signatures of collisionless damping from measurements at only a single point in space. First demonstrated in (gyrokinetic) simulations, this technique was later applied to spacecraft data, providing observational evidence for the key role of electron Landau damping for dissipative processes in space plasmas [43, 44].

Meanwhile, significant recent progress has also been made in terms of a better understanding of the energy partition between electrons and ions during the heating process. It was shown, in particular, that in low-beta, weakly collisional plasmas, such as the solar corona and some instances of the solar wind, Alfvénic fluctuations produce no ion heating within the gyrokinetic approximation [45]. Consequently, all low-frequency ion heating in such plasmas should be due to compressive fluctuations (so-called slow modes). Since these fluctuations energetically decouple from the Alfvénic ones already in the inertial range, this means that the energy partition between ions and electrons in low-beta plasmas is decided at the outer scale, where turbulence is launched, and can be determined from MHD models of the relevant astrophysical systems. Any additional ion heating must come from

mechanisms which are explicitly excluded from gyrokinetics, such as cyclotron heating or the stochastic heating. In a related study, compressively driven turbulence was investigated by means of a modified gyrokinetic model. In this context, a prescription for the ion-to-electron heating ratio was derived, as a function of the compressive-to-Alfvénic driving power ratio, the ratio of ion thermal pressure to magnetic pressure, and the ratio of ion-to-electron background temperatures [46].

These are valuable contributions to solving the turbulent dissipation problem provided by gyrokinetic simulations. They complement and extend work based on fully kinetic and hybrid kinetic-fluid models. An overview summarizing both observational and computational results, including gyrokinetic simulations, was recently put forth by D. Verscharen and colleagues [47].

4.4.1.3 Brief Discussion of the Validity of the Gyrokinetic Simulation Results

As emphasized earlier, gyrokinetics is based on certain assumptions (“gyrokinetic ordering”) like $\omega \ll \Omega$. This rules out, right from the start, the occurrence of certain types of plasma waves, including, in particular, whistlers. Given that whistlers are the second main candidate—besides kinetic Alfvén waves—to explain the observed solar wind fluctuations, this naturally leads to the question: What is the role of whistler turbulence in turbulent dissipation?

To address this important issue head-on, one needs to transcend gyrokinetics and resort to a fully kinetic approach.³ This has been done, for example, in Ref. [48]. Six-dimensional PIC simulations of plasma turbulence for solar wind parameters, which explicitly include whistlers, showed that the spectral properties at kinetic scales are consistent with theoretical expectations for kinetic Alfvén waves and that the contributions of whistlers are subdominant. Similar conclusions have been reached in studies analyzing spacecraft observation data [49, 50]. While these investigations cannot completely rule out a potential role of whistlers under certain circumstances, they clearly indicate that solar wind turbulence at about 1 AU tends to have a strong Alfvénic component. And gyrokinetics is perfectly suited to further its understanding.

4.5 Outlook

Over the last decade or so, gyrokinetics—originally developed in the context of fusion research—gradually found various applications in space plasma physics. Arguably, the most prominent example is the study of solar wind turbulence. Here, gyrokinetics has already made significant contributions, complementing simulation studies based on hybrid kinetic-fluid and fully kinetic models. It can be expected that the existing work will be extended in the future into different parameter regimes,

³ A hybrid kinetic-fluid approach would also shed some light on this issue, but it is hampered by the absence of electron Landau damping.

also addressing turbulence in other types of magnetized space plasmas. Moreover, the physics of magnetic reconnection in the presence of a strong guide field as well as the interaction between turbulence and magnetic reconnection are likely to receive a significant amount of additional attention. Besides all that, gyrokinetic simulation can also serve as a bridge between laboratory experiments and space plasma observations, providing a joint analysis platform. Yet different applications are also bound to emerge. The public availability of gyrokinetic codes like GENE (<http://genecode.org>) is expected to foster a growing number of gyrokinetic studies in space plasma physics.

References

1. G.W. Hammett, F.W. Perkins, Phys. Rev. Lett. **64**, 3019 (1990)
2. G.W. Hammett, W. Dorland, F.W. Perkins, Phys. Fluids B **4**, 2052 (1992)
3. W. Dorland, G.W. Hammett, Phys. Fluids B **5**, 812 (1993)
4. N. Mattror, S.E. Parker, Phys. Rev. Lett. **79**, 3419 (1997)
5. N. Mattror, Phys. Plasmas **5**, 1822 (1998)
6. N. Mattror, Phys. Plasmas **6**, 1065 (1999)
7. A.J. Brizard, T.S. Hahm, Rev. Mod. Phys. **79**, 421 (2007)
8. J.R. Cary, A.J. Brizard, Rev. Mod. Phys. **81**, 693 (2009)
9. J.A. Krommes, Annu. Rev. Fluid Mech. **44**, 175 (2012)
10. G.G. Howes, S.C. Cowley, W. Dorland, G.W. Hammett, E. Quataert, A.A. Schekochihin, Astrophys. J. **651**, 590 (2006)
11. D. Told, J. Cookmeyer, F. Muller, P. Astfalk, F. Jenko, New J. Phys. **18**, 065011 (2016)
12. D. Groselj, S.S. Cerri, A.B. Navarro, C. Willmott, D. Told, N.F. Loureiro, F. Califano, F. Jenko, Astrophys. J. **847**, 28 (2017)
13. E.A. Frieman, L. Chen, Phys. Fluids **25**, 502 (1982)
14. C.D. Stephens, R.W. Brzozowski, F. Jenko, Phys. Plasmas **24**, 102517 (2017)
15. T.G. Northrop, *The Adiabatic Motion of Charged Particles* (Interscience Publishers, New York, 1963)
16. W.W. Lee, Phys. Fluids **26**, 556 (1983)
17. A.Y. Aydemir, Phys. Plasmas **1**, 822 (1994)
18. W.M. Nevins, G.W. Hammett, A.M. Dimits, W. Dorland, D.E. Shumaker, Phys. Plasmas **12**, 122305 (2005)
19. A.M. Dimits, W.W. Lee, J. Comput. Phys. **107**, 309 (1993)
20. S.E. Parker, W.W. Lee, Phys. Fluids B **5**, 77 (1993)
21. R. Hatzky, R. Kleiber, A. Koenies, A. Mishchenko, M. Borchardt, A. Bottino, E. Sonnen-drueler, J. Plasma Phys. **85**, 905850112 (2019)
22. Z. Lin, T.S. Hahm, W.W. Lee, W.M. Tang, R.B. White, Science **281**, 1835 (1998)
23. Y. Idomura, S. Tokuda, Y. Kishimoto, Nucl. Fusion **43**, 234 (2003)
24. W.X. Wang, T.S. Hahm, W.W. Lee, G. Rewoldt, J. Manickam, W.M. Tang, Phys. Plasmas **14**, 072306 (2007)
25. T. Korpilo, A. Gurevich, E. Gusakov, J. Heikkinen, S. Janhunen, T. Kiviniemi, S. Leerink, P. Niskala, A. Perevalov, Comput. Phys. Commun. **203**, 128 (2016)
26. S. Ku, C.S. Chang, R. Hager, R.M. Churchill, G.R. Tynan, I. Cziegler, M. Greenwald, J. Hughes, S.E. Parker, M.F. Adams, E. D'Azevedo, P. Worley, Phys. Plasmas **25**, 056107 (2018)
27. E. Lanti, N. Ohana, N. Tronko et al., Comput. Phys. Commun. **251**, 107072 (2020)
28. F. Jenko, W. Dorland, M. Kotschenreuther, B.N. Rogers, Phys. Plasmas **7**, 1904 (2000)
29. W. Dorland, F. Jenko, M. Kotschenreuther, B.N. Rogers, Phys. Rev. Lett. **85**, 5579 (2000)
30. V. Grandgirard, J. Abiteboul, J. Bigot et al., Comput. Phys. Commun. **207**, 35 (2016)

-
31. J. Candy, R.E. Waltz, *J. Comput. Phys.* **186**, 545 (2003)
 32. T.-H. Watanabe, H. Sugama, *Nucl. Fusion* **46**, 24 (2006)
 33. Y. Idomura, M. Ida, T. Kano, N. Aiba, S. Tokuda, *Comput. Phys. Commun.* **179**, 391 (2008)
 34. A.G. Peeters, Y. Camenen, F.J. Casson et al., *Comput. Phys. Commun.* **180**, 2650 (2009)
 35. A.A. Schekochihin, S.C. Cowley, W. Dorland et al., *Astrophys. J. Suppl. Ser.* **182**, 310 (2009)
 36. R. Bruno, V. Carbone, *Living Rev. Solar Phys.* **10**, 2 (2013)
 37. G.G. Howes, W. Dorland, S.C. Cowley, G.W. Hammett, E. Quataert, A.A. Schekochihin, T. Tatuno, *Phys. Rev. Lett.* **100**, 065004 (2008)
 38. G.G. Howes, J.M. TenBarge, W. Dorland, E. Quataert, A.A. Schekochihin, R. Numata, T. Tatuno, *Phys. Rev. Lett.* **107**, 035004 (2011)
 39. D. Told, F. Jenko, J.M. TenBarge, G.G. Howes, G.W. Hammett, *Phys. Rev. Lett.* **115**, 025003 (2015)
 40. A.B. Navarro, B. Teaca, D. Told, D. Groselj, P. Crandall, F. Jenko, *Phys. Rev. Lett.* **117**, 245101 (2016)
 41. B. Teaca, F. Jenko, D. Told, *New J. Phys.* **19**, 045001 (2017)
 42. B. Teaca, A.B. Navarro, D. Told, T. Gorler, G. Plunk, D.R. Hatch, F. Jenko, *Astrophys. J.* **886**, 65 (2019)
 43. K.G. Klein, G.G. Howes, J.M. TenBarge, *J. Plasma Phys.* **83**, 535830401 (2017)
 44. C.H.K. Chen, K.G. Klein, G.G. Howes et al., *Nat. Comm.* **10**, 740 (2019)
 45. A.A. Schekochihin, Y. Kawazura, M.A. Barnes, *J. Plasma Phys.* **85**, 905850303 (2019)
 46. Y. Kawazura, A.A. Schekochihin, M. Barnes, J.M. TenBarge, Y. Tong, K. G. Klein, W. Dorland, *Phys. Rev. X* **10**, 041050 (2020)
 47. D. Verscharen, K.G. Klein, B.A. Maruca, *Liv. Rev. Sol. Phys.* **16**, 5 (2019)
 48. D. Groselj, A. Mallet, N.F. Loureiro, F. Jenko, *Phys. Rev. Lett.* **120**, 105101 (2018)
 49. C.S. Salem, G.G. Howes, D. Sundkvist, S.D. Bale, C.C. Chaston, C.H.K. Chen, F.S. Mozer, *Astrophys. J.* **745**, L9 (2012)
 50. C.H.K. Chen, S. Boldyrev, Q. Xia, J.C. Perez, *Phys. Rev. Lett.* **110**, 225002 (2013)



Eulerian Approach to Solve the Vlasov Equation and Hybrid-Vlasov Simulations

5

Francesco Califano and Silvio Sergio Cerri

Abstract

In this chapter we give an overview on Eulerian Vlasov algorithms and their applications, including few basic and advanced examples. In fact, thanks to recent advances in high-performance computing, the applications of Vlasov codes have experienced a substantial increase. These applications range from the astrophysical to the laboratory framework through space plasmas. Eulerian Vlasov algorithms are particularly suited for investigating specific kinetic processes because of their very low noise level, especially in the presence of a strong nonlinear dynamics. This is the case, for instance, of nonlinear Landau damping and saturation of electromagnetic instabilities, for which we report to basic examples. Finally, we provide also a brief overview of their advanced applications in the context of kinetic plasma turbulence.

5.1 Introduction

Nowadays, almost all frontier problems in plasma physics must necessarily face the comprehension of (nonlinear) kinetic processes and multi-scale dynamics underlying the evolution of space and laboratory systems. The main reason for this necessity is given by the (almost) lack of collisions in most of such plasmas. Due to the intrinsic nonlinear dynamics of “collisionless” plasmas, even when the energy is

F. Califano (✉)

Department of Physics “E. Fermi”, Università di Pisa, Pisa, Italy
e-mail: francesco.califano@unipi.it

S. S. Cerri

Université Côte d’Azur, Observatoire de la Côte d’Azur, CNRS, Laboratoire Lagrange, France
e-mail: silvio.cerri@oca.eu

initially injected at the large “fluid” scales, the system will, sooner or later, generate smaller and smaller scales, eventually reaching the so-called plasma micro-scales. This means that kinetic physics will have to be included in the picture: typically ion kinetic effects will come into play first and eventually those associated to the electrons. One of the most representative examples of such a behavior is plasma turbulence, in which the underlying process of energy cascade can be considered as the “*rosetta stone*” for the study of nonlinear, multi-scale physics, including the competition between local and non-local energy transfer mechanisms. In fact, the energy cascade process in plasma turbulence is not limited to the classical picture of what is usually referred to as “wave-wave interaction,” being the realm of many fundamental (fluid and kinetic) instabilities, from magnetic reconnection to temperature anisotropy and beam-driven instabilities, as well as by wave-particle interactions. These processes all together contribute to set up a very complex turbulent, collisionless plasma dynamics. Furthermore, these instabilities, which could eventually contribute to the overall energy transfer, do not develop only in the initial phase of turbulence kick-off. Indeed, there is often an interplay and balance of these effects that brings the system toward a self-regulated, fully developed turbulent state during which instabilities continue to be driven by turbulence and simultaneously feed the turbulence itself.

Turbulence in plasmas not only is a striking example of complex nonlinear, multi-scale physics but stands out also as one of the most fascinating (and unsolved) problem in classical physics. Due to such complexity, a computational approach is required in order to shed light on the physics underlying the dynamics of turbulent plasmas. For these reasons, in this chapter we dedicate a specific section also on advanced applications of Vlasov algorithms to the study of plasma turbulence (Sect. 5.6).

Some context and motivation being (briefly) established, let us now discuss in more details what “including kinetic physics” actually means in a collisionless plasma. The basic equation to be solved for the study of collisionless plasma dynamics is the Vlasov equation for each species’ distribution function (see next section, Eq. (5.1)). This equation evolves the distribution function self-consistently coupled with the Maxwell’s equations (viz., the Faraday and Ampere equations, and Gauss’ law for the electric field, while $\nabla \cdot \mathbf{B} = 0$ is typically used as a numerical check), which in turn involve the source terms encoded in the species’ distribution function. However, depending on the physics to be studied, Maxwell equations can be replaced by the Poisson equation in the electrostatic limit or simplified in the low frequency limit where the displacement current is neglected and the current reduces to the curl of the magnetic field. The same holds for the species’ Vlasov equation as, for instance, in the so-called hybrid approximation, in which such equation is solved only for ions, while the electrons’ Vlasov equation is replaced by a fluid model represented by a generalized Ohm’s law for the electric field. In the high-frequency regime instead, one is mainly interested in the electron dynamics and its evolution on a timescale smaller than the ion timescale. Thus, ions can be assumed just as a fixed neutralizing background. Many other models originating by proper reduction of the Vlasov–Maxwell system exist, each one addressing, as

already stated, different physical regimes. Only once the physics to be investigated and the appropriate model equations are established, one can decide the best suited numerical scheme to solve the corresponding equations.

Starting from the pioneering numerical studies in the fluid framework of magneto-hydrodynamic (MHD) models more than 50 years ago (see Chap. 1 of this book), later on kinetic plasma physics started to be attacked numerically using the “particle-in-cell” (hereafter PIC) method (see Chap. 6 this book). This approach makes use of a Lagrangian technique to advance initially selected slices of the distribution function (hereafter d.f.) to solve the Vlasov equation (no particles are advanced in a PIC code!). The so-called Vlasov Eulerian technique instead requires the full sampling of the phase space. Therefore standard hydrodynamic techniques for the calculation of derivatives (here both in space and velocity) such as the well-known spectral methods are in principle not well suited because they are very expensive in CPU time. Indeed, the Eulerian solution of the Vlasov equation implies to deal with a 6D (3D in space, 3V in velocity, viz., 3D-3V) array plus time. Classical finite difference technique attempts, less computationally expensive, have been rapidly abandoned because of the non-conservation of the main invariants and of the generation of spurious electric fields driven by numerical charge imbalance. Only recently, new strategies for using finite differences techniques with improved conservation properties have been proposed for Eulerian Vlasov codes. However, in order to overcome such a difficulty, an approach based on the so-called *splitting scheme* was originally developed [1, 2]. The idea behind such method was to exploit the “multi-advection” character of the Vlasov equation, as will be discussed in more details in Sect. 5.2. In practice, instead of calculating the space and velocity derivatives, one is left to interpolate the distribution function after moving along the (non-relativistic) characteristics. This approach was found to be very efficient for the pioneering study of the high-frequency electron-scale regime, from λ_D to several tens of λ_D or to Q_e , d_e in the subsequent electromagnetic (hereafter e.m.) approach later on. Here λ_D , Q_e , and d_e are the Debye length, the electron Larmor radius, and the electron skin depth, respectively. In these early studies, typically, kinetic electrons are coupled to the Poisson equation, while ions are in general just considered as a fixed neutralizing background.

In many situations, it is required not only to make use of a numerical algorithm able to follow very accurately resonant wave-particle interactions but also to resolve with high accuracy the small-scale physics, such as electric field fluctuations at (and possibly below) the electron scale within a reconnection layer or low-amplitude fluctuations, as for the tail of a turbulent spectrum well below the ion scale. In these cases, an Eulerian approach (hereafter Vlasov code) is particularly well suited because of their almost “noise-free” algorithms, provided that processes such as particle acceleration or formation of shock structures are not too extreme since the velocity space grid is fixed at the beginning of the simulation. Otherwise a (semi) Lagrangian PIC scheme should be adopted. Note that the Eulerian approach is typically more demanding in terms of computational resources because the full phase space must be properly sampled from the beginning of a simulation. Vlasov

and PIC codes are therefore complementary, and the choice between one and the other approach mainly depends on the physics to be investigated.

In the following we will specifically address the Eulerian approach to the numerical solution of the Vlasov equation.

5.2 Numerical Schemes and Hamiltonian Dynamics

The Vlasov equation is the model equation for the study of collisionless magnetized plasma dynamics. The idea is that particles can be represented by a statistical approach and that each charged particle mainly interacts with the full system more than with nearby single particles. In other words, the force acting on the plasma represented statistically by a distribution function is given by the “averaged” electromagnetic field generated by the plasma itself plus external fields rather than by single particle interaction (the so-called mean field theory).

We define the N-particle distribution and the s-particle distribution in the (\mathbf{x}, \mathbf{v}) phase space as

$$F(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N)$$

$$F^{(s)} = \int F d\mathbf{x}_{s+1} \dots d\mathbf{x}_N d\mathbf{v}_{s+1} \dots d\mathbf{v}_N$$

The F distribution function satisfies the Liouville equation,

$$\frac{dF}{dt} = 0$$

where d/dt is the total derivative given by

$$\frac{d}{dt} \equiv \frac{\partial}{\partial t} + \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{x}} + \mathbf{a} \cdot \frac{\partial}{\partial \mathbf{v}}$$

Here \mathbf{a} represents the total acceleration due to internal and external electromagnetic fields. Then, formally, the Vlasov equation is obtained by integrating the Liouville equation for the N-particle distribution function over all particles but one (see [3] for the full theory)

$$\int \frac{d}{dt} F(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N) d\mathbf{x}_2 \dots d\mathbf{x}_N d\mathbf{v}_2 \dots d\mathbf{v}_N \rightarrow \frac{dF^{(1)}}{dt} = 0$$

neglecting the particle correlations at any level, $F^{(2)} = 0$, and so on. This means that collisions are completely neglected, as consistent with the large majority of plasma systems in space, let's say border line in the laboratory. For instance, in the solar wind where non-Maxwellian d.f. is routinely detected by satellite measurements,

the mean free path is of the order of one astronomical unit corresponding to the typical size of the system. So, neglecting the two-particle correlation function and so on, the resulting equation for the one point d.f. turns out to be a “multi-advection” equation in phase space satisfying the Liouville theorem, $df/dt = 0$ (from now on $F^{(1)} \equiv f$). This equation is known as the Vlasov equation, and for a given species, α , in the non-relativistic limit, reads:

$$\frac{\partial f_\alpha}{\partial t} + \mathbf{v} \cdot \frac{\partial f_\alpha}{\partial \mathbf{x}} + \frac{q_\alpha}{m_\alpha} \left(\mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}}{c} \right) \cdot \frac{\partial f_\alpha}{\partial \mathbf{v}} = 0 \quad (5.1)$$

Here the index $\alpha = i, e$ indicates the species (ions and electrons, respectively), m_α and q_α the mass and the electric charge, \mathbf{E} and \mathbf{B} the electric and magnetic field and f_α the distribution function, hereafter d.f. The distribution function is (multi)-adverted in phase space by the space and velocity terms in the sense that any 2D slice (for the sake of simplicity) of the d.f. in phase space can be deformed during the evolution, but its total surface cannot be changed, as schematically represented in Fig. 5.1. Furthermore, by integrating Eq. (5.1), it is easily shown that there are an infinite number of invariants I conserved by the Vlasov equation, in particular any power of the d.f.:

$$I_\alpha^n = \int f_\alpha^n(\mathbf{x}, \mathbf{v}, t) d\mathbf{x} d\mathbf{v} = \text{cost.} \quad (5.2)$$

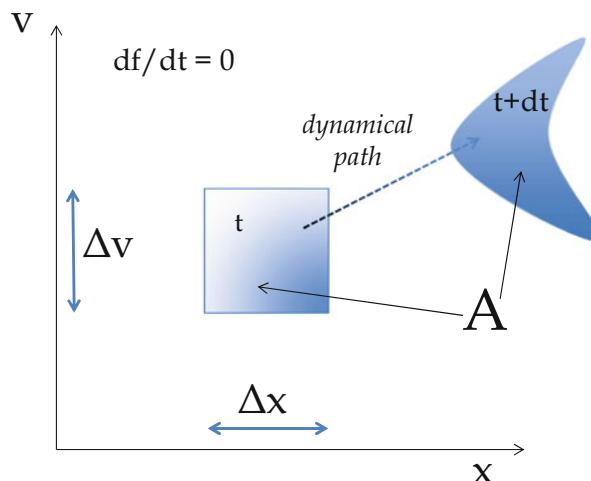


Fig. 5.1 Phase space advection of the distribution function by the Vlasov equation. The phase space shape can be deformed, but its surface (volume) cannot be changed during the dynamical evolution of the plasma

More generally, any function $G(f)$ of f is a constant of motion,

$$\frac{dG}{dt} = \frac{dG}{df} \frac{df}{dt} = 0$$

To evolve in time the Vlasov equation, one must calculate self-consistently the electromagnetic fields using the plasma density and currents obtained from the zero and first moment of f defined for the species α as

$$n_\alpha = \int f_\alpha d\mathbf{v}_\alpha; \quad j_\alpha = q_\alpha \int \mathbf{v} f_\alpha d\mathbf{v}_\alpha \quad (5.3)$$

This is done by solving the Maxwell equations, namely, the Faraday and Ampere equations

$$\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E} \quad (5.4)$$

$$\frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{B} - \frac{4\pi}{c} \mathbf{j} \quad (5.5)$$

or the Gauss equation (or Ampere with $\mathbf{B} = 0$ for non-periodic boundary conditions) in the electrostatic limit:

$$\nabla \cdot \mathbf{E} = 4\pi \int (f_i - f_e) d\mathbf{v} \quad (5.6)$$

The Vlasov equation, Eq. (5.1), is basically a “transport” equation in phase space, and we can distinguish the space advection term \mathcal{A}_x and the velocity advection term \mathcal{A}_v separately (from now on the species index will be ignored):

$$\frac{\partial f}{\partial t} = \mathcal{A}[f] \equiv \mathcal{A}_x[f] + \mathcal{A}_v[f] \quad (5.7)$$

Note that the linear Vlasov equation becomes nonlinear when coupled with Maxwell equations since the velocity advection term \mathcal{A}_v depends on the plasma density and currents and so on the d.f.; see Eq. (5.3), used to solve the Maxwell equations and so the electromagnetic fields. Equation (5.1) has the following characteristics:

$$\frac{ds}{dt} = \mathbf{v}; \quad \frac{d\mathbf{v}}{dt} = \mathbf{a} \quad (5.8)$$

which describe an “Hamiltonian flow” in phase space. Let’s now consider for the sake of mathematical simplicity the electrostatic limit and adopt the “classical” Hamiltonian formulation:

$$\frac{\partial H}{\partial \mathbf{p}} = \frac{\mathbf{p}}{m}; \quad \frac{\partial H}{\partial \mathbf{q}} = -m\mathbf{a} \quad (5.9)$$

where \mathbf{q} and \mathbf{p} are the usual canonical variables, $d\mathbf{q}/dt = \mathbf{p}/m$, and $d\mathbf{p}/dt = m\mathbf{a}$. Here H is the Hamiltonian defined as

$$H = \frac{p^2}{2m} - \Phi(q) \quad (5.10)$$

where $\Phi(q)$ is the electric potential. Using the Hamiltonian, we can rewrite the Vlasov equation as

$$\frac{\partial f}{\partial t} = [H, f]; \quad [H, f] = \frac{\partial H}{\partial x} \frac{\partial f}{\partial v} - \frac{\partial H}{\partial v} \frac{\partial f}{\partial x} \quad (5.11)$$

where $[H, f] = (\partial H / \partial x)(\partial f / \partial v) - (\partial H / \partial v)(\partial f / \partial x) \equiv \Lambda f$ is the Poisson bracket. It allows us to define $\mathcal{P}^t = e^{\Lambda t}$ as the Vlasov propagator whose analytical solution is in general unknown, so that the advection of the d.f. in phase space is purely formal. However, the Hamiltonian can be split into two parts,

$$H = H_s + H_v; \quad H_s = p^2/2m \quad H_v = -\Phi$$

allowing us to define the space and velocity propagator separately

$$\mathcal{P}_s^t = \exp(\Lambda_1 t); \quad \mathcal{P}_v^t = \exp(\Lambda_2 t)$$

The corresponding space and velocity phase space advection

$$\frac{\partial f}{\partial t} = [H_1, f] = -v \frac{\partial f}{\partial x}; \quad \frac{\partial f}{\partial t} = [H_2, f] = E \frac{\partial f}{\partial v}$$

(using now dimensionless variables) can be now solved analytically with solution given by:

$$\mathcal{P}_x^t [f_0] \equiv f(x, v, t) = f_0(x - vt, v); \quad \mathcal{P}_v^t [f_0] \equiv f(x, v, t) = f_0(x, v + Et)$$

In other words, the solution of each advection equation is given by the solution at the previous time evolved along its own characteristic. This method is known as the *splitting scheme* and was proposed many decades ago by Cheng and Knorr [1] for the electrostatic Vlasov equation. However, applying sequentially the space and velocity propagators, $\mathcal{P}_v^t [\mathcal{P}_x^t [f_0]]$, does not correspond to the full advection of the d.f., $\mathcal{P}^t [f_0]$, since the space and velocity propagators do not commute. To solve

the full advection, one should iterate the two propagators, the so-called Magnus expansion [4], as follows:

$$\mathcal{P}^t = \lim_{N \rightarrow \infty} \left[\mathcal{P}_s^{t/2N} \mathcal{P}_v^{t/N} \mathcal{P}_s^{t/2N} \right]^N$$

For a finite time step dt and taking $N = 1$, the solution for the full Vlasov propagator is approximated at second order by:

$$\mathcal{P}^{dt} [f(x, v)] = \left\{ \mathcal{P}_x^{dt/2} \mathcal{P}_v^{dt} \mathcal{P}_x^{dt/2} \right\} [f(x, v)] + O(dt^3) \quad (5.12)$$

provided that the electric field is calculated after the first space translation as discussed in [1].

In the electromagnetic case, one must take into account also the rotation of the d.f. around the magnetic field driven by the Lorentz force, $[E_x + v_y B_z] \partial f / \partial v_x$ and $[E_y - v_x B_z] \partial f / \partial v_y$ (in the 2D limit in space for the sake of mathematical simplicity). In the non-relativistic limit, since each component of the Lorentz force does not depend on the velocity along that component, one can still adopt the same strategy using the corresponding Vlasov propagators

$$\mathcal{P}_v^{dt} = \mathcal{P}_{v_x}^{dt/2} \mathcal{P}_{v_y}^{dt} \mathcal{P}_{v_x}^{dt/2}$$

which is easily extended to the full electromagnetic 3D case [5] as

$$\mathcal{P}_v^{dt} = \mathcal{P}_{v_x}^{dt/2} \left\{ \mathcal{P}_{v_y}^{dt/2} \mathcal{P}_{v_z}^{dt} \mathcal{P}_{v_y}^{dt/2} \right\} \mathcal{P}_{v_x}^{dt/2} \quad (5.13)$$

All this splitting technique is motivated by the need of avoiding direct calculation of phase space derivatives which would be too expensive in terms of CPU time. For the sake of simplicity in the following, let's take the electrostatic 1D-1V limit, but valid in general. By advancing each single advection along its own characteristic, the value of the d.f. at the new time step is obtained by a direct interpolation of the d.f. at the previous time. This can be easily understood by noticing that the new d.f. is given by the value of the d.f. at the previous time shifted by the corresponding advection, for instance,

$$f_s(x, v, t + dt/2) = f_v(x - vdt, v, t)$$

$$E = -\nabla^2 \phi = 4\pi e(n_i - n_e); \quad n_\alpha = \int f_{\alpha,s} dv$$

$$f_v(x, v, t + dt) = f_s(x, v + E dt, t + dt/2) \quad (5.14)$$

where f_s and f_v are the d.f. advected by the space and velocity term, respectively, which are shifted by a $dt/2$ one to each other (see Eq. (5.14)), but which are both advanced by a dt time step.¹ The value of the d.f. in any given phase space grid point at the new time step, $f_a(t + dt, x_i, v_j)$, $a = s, v$, is given by the value of the d.f. at the previous time in the point shifted by the corresponding advection, $\tilde{x} = x_i - vdt$ or $\tilde{v} = E - v_j dt$, respectively. As matter of fact, \tilde{x} and \tilde{v} may not coincide with an actual point of the phase-space grid. Because of numerical stability (see [5]), the time step dt must be “small enough” such that $x_{i-1} < \tilde{x} < x_{i+1}$ and the same holds for the velocity advection. As a result, the splitting scheme reduces to an interpolation of the d.f. in the point \tilde{x} (or \tilde{v}) using the known values in the nearby grid points, x_{i-1}, x_i, x_{i+1} (the stencil can be extended to increase the accuracy).

In the Vlasov code (and its more recent Hybrid-Vlasov version [6]) developed in Pisa in these last years, we use the third-order Van Leer scheme [5] involving a five-point stencil. This interpolation scheme has the advantage of being very accurate, to conserve the charge and to be well suited for parallel machines since it requires only communications between nearby processors.

5.3 Models of Different Plasma Regimes

The great complexity and difficulty in the study of the dynamics of collisionless plasmas stem from the ability of a plasma in supporting an impressive variety of electrostatic and electromagnetic fluctuations at different scales and frequencies. These fluctuations are driven by several different basic physical processes developing at large, intermediate, and small scales where the system is regulated by fluid, ion kinetic, and electron kinetic dynamics, respectively. As discussed in Sect. 5.1, the model equation able to capture all scales except the collisional one is the Vlasov equation, Eq. (5.1), for each species at play self-consistently coupled with Maxwell equations, Eqs. (5.4)–(5.5). The Vlasov equation follows in time the evolution of the d.f. in the six-dimensional 3D-3V phase space. At the level of linear fluctuations underlying any important physical plasma process (turbulence, reconnection, wave-particle interaction, nonlinear interactions, etc.), many modes are at play which develop from fluid down to ion and electron scales: from slow and fast MHD and Alfvén waves to ion acoustic and Bernstein modes, to whistler and kinetic Alfvén waves, and to Langmuir waves (also known as plasma waves). These modes are separated by many orders of magnitudes in wavelength and frequency. As an example, in the solar wind at 1 AU from the sun, Alfvén waves typically vary in the range $10^{-3}\text{--}10^{-1}\text{ Hz}$, while KAW and whistler waves from a bit less than the ion cyclotron frequency, $f_{ci} \approx 1\text{ Hz}$, up to $10\text{--}10^2\text{ Hz}$ not forgetting about ion-sound waves from a few Hz to a few 100 Hz. At even larger frequencies around 10^3 Hz

¹ Technically, a $dt/2$ shift is made at $t = 0$ by a space $dt/2$ advection. Then, the iteration of the last and first $dt/2$ space advection is cast into a single dt space advection. To obtain the “correct” d.f. $f(x, v, t)$ at a given time, if needed, one must shift by $dt/2$ in space the f_v d.f.

and more, one observes plasma waves. Linear modes in a fusion machine, from TAE (Toroidal Alfvén Eigenmodes) and ITG (Ion Temperature Gradient)-driven modes to electrostatic and drift modes, are observed in a less extended range of frequencies, but still spanning a quite large range of at least three orders of magnitude, from $k\text{Hz}$ to $M\text{Hz}$. Turbulence development extends further the range of frequencies involved in the full system dynamics.

Wave modes introduced above are just the most popular ones and among the most representative of the different regimes shown in Fig. 5.2. However, in general even starting in the context of a well-defined regime, the system dynamics extend, very rapidly in some cases, to nearby regimes. In general such multi-scale behavior is limited to very localized regions, e.g., current sheets, but nevertheless the consequences can have a global impact (magnetic field topology change by reconnection, particle acceleration, etc.). In these conditions other linear modes come into play depending on the processes at play that feed such regime transition self-consistently achieved by the system. Indeed, plasmas are known as multi-scale, multi-physics systems. A similar sketch as for Fig. 5.2 of many different instabilities supported by a plasma depending on the regime of interest could be drawn. Note that instabilities are among the main drivers for regime transition. As a consequence, addressing the dynamics of a magnetized plasma by numerical simulations not only requires very accurate algorithms to describe the nonlinear dynamics, but it also imposes to deal with a very large interval in wave numbers and frequencies corresponding to the requirement of a huge number of grid points both in space and velocity and a very large number of time steps. As an example, last year's Vlasov-hybrid kinetic simulations can achieve up to tens of Tb of memory and running time

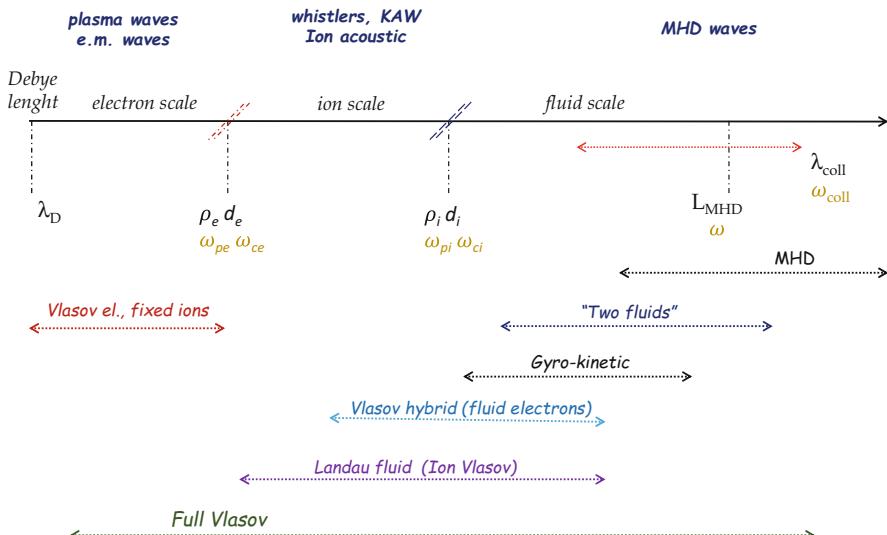


Fig. 5.2 The different theoretical models used in plasma physics at different scales

of the order of tens of millions of core hours, which is at the limit of the present machines. A related technical problem to such a huge use of numerical resources, but nevertheless of strong consequences on the possibility to study the physics of interest, is the data storage and analysis in particular of the d.f.

Therefore, the solution of the full Vlasov–Maxwell system necessitates of an enormous investment in terms of computational resources both of CPU time and memory. It turns out that even using the most powerful supercomputers it is in practice impossible to fully describe phenomena such as the interaction of the solar wind with the magnetosphere or the development of turbulent fluctuations in the solar wind where at least four orders of magnitudes of scales and frequencies are involved. For instance, in the solar wind, the ratio between the plasma frequency and the ion cyclotron frequency is about 10^3 as shown in [7] (see Fig. 2), which means that more than four orders of magnitudes of frequencies should be taken into account in a Vlasov simulation if one wants to describe the full system, from MHD up to electron frequencies. Correspondingly, a likewise very large interval in wave vectors must be taken into account.

For all these reasons, it is very important to separate the different actors at play in the full scenario of the system dynamics in order to adopt specific models to investigate selected pieces of physics. For example, global simulations of the interaction of the solar wind–Earth’s magnetosphere using plasma parameters as given by satellite observations are still performed mainly using an MHD approach (see also Chap. 1 of this book), while to investigate the physics developing in the region of interaction, from the magnetosheath to the external magnetosphere, one can make use of hybrid models to follow the crossing from fluid to ion kinetic dynamics. A typical example is the study of the low latitude dynamics driven by the development of Kelvin-Helmholtz vortices initially on a fluid scale but rapidly exciting a dynamics across d_i and/or ϱ_i (the ion skin depth and the ion Larmor radius, respectively) down to sub-ion’s scales.

One of the most popular models to investigate the physics of the low-frequency MHD dynamics and its transition to the ion kinetic regime is the so-called Hybrid approach where the ion kinetic dynamics is advanced in time by the Vlasov equation while the electrons are advanced using a fluid approach [8]. Quasi-neutrality is also assumed. Hybrid approach, long time ago developed within the PIC framework (see Chap. 3 of this book), is motivated by the need of following the fluid to ion-kinetic crossing of the dynamics but avoiding to take into account the electron kinetic dynamics too because of computational reasons. This is in part justified by the low frequency physics mainly at play in many of the space systems and the possibility of a transition across the ion kinetic characteristic length scale because of nonlinear interactions and/or instabilities development driven by plasma temperature anisotropy and/or by gradients at the interface between different systems or because of turbulence. Nevertheless, the lack of the electron kinetic dynamics remains a weak point because of the multi-scale behavior of plasma (with possible feedback from small to large scales) and because of the difficulties to

impose a fluid closure² (typically isothermal) which could be not at all appropriate once the dynamics develop. On the top of that, electron Landau damping, lacking in the hybrid approach, can also be an important piece of physics first of all because it influences the dispersion relation of the main wave modes at frequencies overcoming the ion cyclotron frequency. Saying that, the hybrid approach has been (and it is still) a very powerful approach to investigate fundamental problems, for instance, the dynamics of collisionless shocks, magnetic reconnection, temperature anisotropy instabilities, and plasma turbulence across the ion cyclotron frequency, in all cases being able to catch the main ion kinetic physics of interest.

We start by considering the Vlasov-Eulerian approach known as the HVM (Hybrid Vlasov Maxwell) code presented long time ago in [6]. The numerical scheme to advance the ion Vlasov equation is based on the splitting scheme (see Sect. 5.2) coupled to a discontinuous Galerkin method (see [5] for details). The hybrid scheme can couple the ion Vlasov equation to fluid electrons, e.g., by means of a current advance method and cycling leapfrog (CAM-CL) approach [9]. The CAM-CL method has been extensively adopted by Hybrid PIC algorithms since it requires only a single computational step (i.e., only one integration of the d.f. to calculate the plasma density and current) to complete the full advancement. This is achieved by advancing the ionic current using an appropriate equation of motion such as to maintain second-order time accuracy.

A different approach with respect to the splitting scheme has been proposed more by Elkina and Büchner [10]. The main goal was to ensure the positivity of the d.f., thus avoiding numerical artificial oscillations that could lead to code instabilities. The time advancing scheme is given by a second-order upwing algorithm together with a finite volume approach for grid discretization.

In these last years, a new 3D-3V hybrid-Vlasov code, namely, Vlasiator (see [11] and references therein), aimed at performing global magnetospheric simulations [12, 13], has been developed. This approach was motivated by the ability of the MHD model of capturing the global dynamics of the solar wind-magnetosphere system but at the same time by its inability to follow the kinetic dynamics developing first of all at the interfaces of different plasma regions. It has been the first Vlasov code developed to study the near Earth's environment. It integrates the Vlasov equation for the ion population by means of a finite volumes technique and adopting a splitting scheme similar to what has been presented for the HVM code in Sect. 5.2; see Eq. (5.12). The electrons are instead considered as a massless neutralizing background. The system is closed by using an electric field calculated by an ideal Ohm's law in the Faraday equation, while the electric field in the Vlasov

² Formally, fluid equations are recovered by multiplying the Vlasov equation by \mathbf{v}^n , $n = 0, 1, \dots$, and integrating in the velocity \mathbf{v} space (the so-called n-moments). This leads to a “useless” infinite chain of equations (see, e.g., [3]). The fluid closure is given by a supplementary equation. In general, one assumes an isotropic pressure and takes a polytropic relation between density and pressure eventually breaking the chain and closing the system of equations. However, this supplementary equation, initially motivated by the expected physics, can rapidly uncorrectly describe the dynamics at play.

equation contains also the Hall term. Finally quasi-neutrality is assumed, and the displacement current is neglected in the Ampere law. A dynamical time step is used to adjust it to the maximum allowed value depending on the dynamics. Vlasiator approach is very well suited for performing global magnetospheric simulations including boundaries where the kinetic physics at ion scales dominate the dynamics, for instance, the Earth's bow shock. The only compromise, as for all hybrid codes, is the neglecting of the electron scale kinetic processes, but being able to follow the multi-scale plasma physics on a large number of scales.

As of today, the full kinetic Vlasov approach proposed long time ago by Mangeney and Califano [5] has not been exploited yet unless, for instance, in reduced geometry, for instance, 1D-3V geometry for the study of electromagnetic instabilities (see [14, 15] and references therein). This is the direct consequence of the very huge CPU time and memory requirement of the Eulerian approach where the full grid in phase space must be set from the initial time. For this reason the hybrid approach, 2D-3V first and more recently 3D-3V, has been so far so popular. However the impressive increase in computational power in the last recent years (the so-called Peta-scale computing and soon achieving Exa-scale) now finally allows for fully kinetic Vlasov simulations. Recently, a new algorithm for the solution of the Vlasov Maxwell system still in the non-relativistic limit has been presented [16]. This approach is based on a discontinuous Galerkin technique and makes use of finite element method for the spatial discretization. Moreover, in this approach the Vlasov equation is integrated directly without using the splitting technique for the two advection terms. The explicit time advancement, namely, a third-order Runge-Kutta scheme, is constrained by the CFL condition (For more details about the CFL condition, see Chap. 6, Sect. 6.2.1 “Field Solvers”) on the time step resulting from the discretization of the two advection terms. Despite the highly sophisticated parallelization technique, the use of a full kinetic Eulerian Vlasov code still remains challenging when applied to research problems. Very recently, still on the line of fully kinetic simulations, a full Vlasov numerical code adopting the Darwin approximation has been developed [17]. This code, namely, ViDA, makes use of the Vlasov-Darwin model as the one adopted in [18]. The advancement of the distribution function is the same as the one discussed for the Hybrid-Vlasov (HVM) code introduced above. However, instead of coupling the d.f. advancement to the full set of Maxwell equations, it relies on the Darwin approximation where light waves are suppressed while standard wave modes are included (e.g., kinetic Alfvén, whistler waves, etc.). In particular, the standard hyperbolic time evolving Faraday and Ampere equations can be cast into a set of Helmholtz- and Poisson-like equations. This approach allows one to significantly increase the time step with respect to a Vlasov–Maxwell model since light waves are in general the fastest process to follow within the framework of an explicit algorithm. As a consequence, this approach is particularly adapted for the study of the dynamics in between the ion and the electron kinetic scale length, i.e., in between q_i, d_i , and q_e, d_e (the ion/electron Larmor radius and the ion/electron skin depth, i/e, respectively). This kind of approach aims at investigating multi-scale plasma problems where there is a need of a fine, no noise resolution of the electron physics embedded into an ion

large-scale dynamics. Important examples are the study of kinetic turbulence and “dissipative”-like properties or of the structure of the ion/electron reconnection layer for direct comparison with satellite data such as those from NASA *Magnetospheric Multi-Scale* (MMS) mission (the so-called electron diffusion layer).

5.4 Vlasov Equilibrium

Any study aimed at investigating the dynamics of a specific plasma process, wave propagation, instabilities up to the nonlinear dynamics or the development of plasma turbulence should mandatory start by finding an initial equilibrium configuration represented by the d.f. and the fields, $f_\alpha(\mathbf{r}, \mathbf{v})$, $\mathbf{E}(\mathbf{r})$, and $\mathbf{B}(\mathbf{r})$. This initial configuration must satisfy the Vlasov equation, Eq. (5.1), self-consistently coupled to the Maxwell equations, Eqs. (5.4)–(5.5) in the stationary limit $\partial/\partial t = 0$. This can be achieved by using the “integrals of motion” starting from the single particle Hamiltonian:

$$\mathcal{H}_\alpha = p_\alpha^2/2m_\alpha + q_\alpha\phi_\alpha(\mathbf{r}); \quad \mathbf{p}_\alpha = m_\alpha\mathbf{v}_\alpha + q_\alpha\mathbf{A}/c \quad (5.15)$$

where $\phi(\mathbf{r})$ and $\mathbf{A}(\mathbf{r})$ are the potentials satisfying the following equations

$$\nabla^2\mathbf{A} = -\frac{4\pi}{c}\mathbf{J}(\mathbf{A}, \phi); \quad \nabla^2\phi = -4\pi\rho(A, \phi) \quad (5.16)$$

In the space plasma context, collisions act on a timescale which is in general much longer than the dynamical timescale, a condition which remains more or less true also in the laboratory (except that for special cases). Therefore plasmas are typically non-equilibrium systems, and Vlasov equilibria are thus defined as meta-equilibria. As an example, anisotropic distributions such as

$$f(\mathbf{v}) = C \exp \left[-\frac{1}{2m} \left(\frac{v_x^2}{T_x} + \frac{v_y^2}{T_y} + \frac{v_z^2}{T_z} \right) \right] \quad (5.17)$$

are formally solutions of the stationary Vlasov equation.

The most popular equilibrium because of its simplicity in modeling the reconnection problem is the Harris sheet equilibrium [19]. It is a 1D model equilibrium based on the assumption of no charge separation, $\nabla^2\phi = 0$, giving $\phi = cts$. Then the integrals of motion are the canonical momentum, the energy, and the component of the velocity along the magnetic field direction. Assuming x the coordinate of variation and y the direction of the magnetic field, we take $f = f(p_z, v_y, v^2)$ and $\mathbf{A} = A_z(x)\mathbf{e}_z$, so that one should solve the following equation:

$$\frac{d^2A_z}{dx^2} = -\frac{4\pi e}{c} \int (f_i - f_e) v_z d^3v \quad (5.18)$$

Here \mathbf{A} is the magnetic vector potential defined as $\nabla \times \mathbf{A} = \mathbf{B}$. By using a Hamiltonian formulation where the force term is given by the current in the z -direction, one can define an equivalent effective potential $\mathcal{D}(A_z) = \int_{A_0}^{A_z} J_z(A'_z) dA'_z$ and get an Energy equivalent equation:

$$\frac{1}{2} (A'_z)^2 + \mathcal{D}(A_z) = \text{cost.} \quad (5.19)$$

The effective potential can be interpreted by considering the force balance equation, $(1/c) \mathbf{J} \times \mathbf{B} = \nabla \cdot \mathbf{P}$. Still assuming $\mathbf{A} = A_z(x)\mathbf{e}_z$, one gets an energy-like equation [20]

$$\frac{d}{dx} \left[\frac{1}{2} \left(\frac{dA_z}{dx} \right)^2 + 4\pi P_{xx} \right] = 0 \quad (5.20)$$

which, by comparing with Eq. (5.19), finally gives the effective potential

$$\mathcal{D}(A_z) = 4\pi P_{xx}(A_z)$$

As a result, the equilibrium configuration is not imposed but results from the solution of Eq. (5.20). In other words the vector potential is obtained, let's say, in a blinded way, and thus the magnetic field configuration can be in principle not well suited for the problem of interest. Furthermore, one must check *a posteriori* that quasi-neutrality is satisfied and that $f(p_z, v_y, v^2)$ is positive definite.

The most important model of the above theory is the well-known Harris equilibrium, a magnetic sheared 1D configuration largely used in kinetic simulations aimed at investigating the magnetic reconnection process. In this case the d.f. of both species can be written as:

$$f_\alpha = f_0 N_H \exp \left\{ -\frac{m_\alpha}{2T_\alpha} \left[v_{x,\alpha}^2 + v_{y,\alpha}^2 + (v_{z,\alpha} - u_{*,\alpha})^2 + \frac{2u_{*,\alpha}q_\alpha}{m_\alpha c} A_z \right] \right\} \quad (5.21)$$

where $u_{*,\alpha}$ is the electron and ion mean velocity, $\alpha = i, e$, respectively, $f_0 = (m_\alpha/2\pi T_\alpha)^{3/2}$, and N_H is a normalization factor. By combining Eq. (5.21), easily rewritten in terms of the integral of motion, with Maxwell equations

$$\nabla \cdot \mathbf{E} = 4\pi \Sigma_\alpha q_\alpha \int f_\alpha du_\alpha^3$$

$$\nabla \times \mathbf{B} = (4\pi/c) \Sigma_\alpha q_\alpha \int f_\alpha \mathbf{u}_\alpha du_\alpha^3$$

one gets the equations for the potential:

$$\frac{d^2\varphi}{dx^2} = 4\pi e N_H \left\{ \exp \left[\frac{eu_{*,e}}{cT_e} A_z - \frac{e}{T_e} \varphi \right] - \exp \left[-\frac{eu_{*,i}}{cT_i} A_z + \frac{e}{T_i} \varphi \right] \right\} \quad (5.22)$$

$$\frac{d^2A_z}{dx^2} = -\frac{4\pi e N_H}{c} \left\{ \frac{u_{*,e}}{T_e} \exp \left[\frac{eu_{*,e}}{cT_e} A_z - \frac{e}{T_e} \varphi \right] - \frac{u_{*,i}}{T_i} \exp \left[-\frac{eu_{*,i}}{cT_i} A_z + \frac{e}{T_i} \varphi \right] \right\} \quad (5.23)$$

To get the Harris solution, one must impose the quasi-neutrality condition, $\phi = 0$. Thus from Eq. (5.22) one gets the constraint

$$\frac{u_{*,i}}{T_i} + \frac{u_{*,e}}{T_e} = 0 \quad (5.24)$$

By using the Hamiltonian formulation using the integrals of motion and assuming the force term given by the current in the z -direction, one gets the effective potential $\mathcal{D}(A_z) = \mathcal{D}_0 \exp[A_z/\bar{A}_z]$. Finally, we integrate Eq. (5.19) and get the following solution

$$A_z(x) = -\frac{2cT_e}{eu_{*,e}} \log [\cosh(x/L)] ; \quad \frac{1}{L} = \frac{eu_{*,i}}{cT_i} \sqrt{2\pi N_H(T_i + T_e)} \\ B_y(x) = \frac{eu_{*,e}}{cT_e} \sqrt{2\pi N_H(T_e + T_i)} \tanh(x/L) \quad (5.25)$$

Assuming $T_e \simeq T_i$, the equilibrium scale length can be written as $L \simeq d_e(v_{th,e}/u_{*,e})$ so that if the characteristic velocity supporting the current is much smaller than the thermal velocity, $u_{*,e} \ll v_{th,e}$, we get $L \gg d_e$, a characteristic condition corresponding to the MHD regime. Here $d_e = c/\omega_{pe}$ indicates the electron skin depth. Finally, by integrating the d.f., one obtains the density distribution as follows:

$$n_i(x) = n_e(x) = n_H \cosh^{-2}(x/L) \quad (5.26)$$

This equation shows that the density far from the central current layer goes to zero which is often an unrealistic condition, in particular for what concerns space plasmas. Therefore, the Harris sheet model often includes a constant supporting density, a so-called platform, given by a Maxwellian which is a function of the energy only and thus still represents an equilibrium:

$$F_\alpha(x) = f_{\alpha H}(x) + f_0 N_p \exp \left\{ -\frac{m_\alpha}{2T_\alpha} \left[u_{x,\alpha}^2 + u_{y,\alpha}^2 + u_{z,\alpha}^2 \right] \right\} \quad (5.27)$$

The Harris equilibrium plus platform can be simply viewed as a current layer, particularly suited for the study of magnetic reconnection [21, 22] because of its capability of capturing the main physics of reconnection in spite of its technical and mathematical simplicity, a feature particularly useful for numerical simulations.

In general, because of computational reasons, periodic boundary conditions are used, a constraint not suited for the Harris equilibrium. To circumvent this inconvenient, in many studies a double Harris sheet is employed, that is, a simple sum of two opposite Harris equilibria so that the magnetic field profile can now be assumed as periodic. However, the so-called double-Harris sheet is no more a kinetic equilibrium because the first of Eq.(5.16) is a nonlinear equation and thus not satisfied by a simple linear combination of Harris sheets. Nevertheless, the double-Harris sheet is a force balance (MHD) equilibrium and thus has been deeply used in kinetic simulations by assuming that its characteristic time to evolve is longer than the system dynamics timescale or that it rapidly relaxes to a nearby kinetic equilibrium. However, this point remains since the above assumptions are difficult to be proven. More generally, even for a stationary Vlasov equilibrium, one must check the stability against phase space instabilities in the velocity space, for example, those due to the presence of local beam-like configurations or temperature anisotropy that would partially or even completely mask the dynamics of interest because, for example, of the impact on the growth rate of the reconnection instability [23, 24].

An analytical solution for a “double-Harris like configuration” has been obtained by superposing two opposite currents corresponding to two counter-streaming electron populations leading to a sinusoidal magnetic equilibrium but involving temperature anisotropic distributions; see [25]. These Harris-like approaches still face with the problem of obtaining an equilibrium solution in a “blinded way,” so that the final equilibrium is in principle “unknown” at the beginning of the procedure. In other words at the end of the procedure, the equilibrium solution could be far from the one to be used. To step over this problem, a generalized method for the “inverse problem” in 1D has been developed and summarized in [26]. It is aimed at obtaining analytical Vlasov equilibrium supporting a given macroscopic current sheet, including the case of force-free equilibria. The main objective is to address the magnetic reconnection problem. In this case, still using an Hamiltonian formulation in terms of the canonical momenta $p_\alpha = mv_\alpha + qA_\alpha/c$ and the corresponding Hamiltonian $H = mv^2/2 + q\phi$ for each species, one starts by imposing the force balance between the Lorentz force and the pressure term using the standard definition for the species pressure tensor:

$$\Pi_{ij} = \sum_s m \int (v_i - u_{*,i})(v_j - u_{*,j}) f dv^3 \quad (5.28)$$

where $u_{*,i/j}$ is the i/j component of the mean velocity and the species index α has been omitted. The idea is to write the d.f. of each species in Eq.(5.28) in terms of a subset of constant of motions multiplied by an unknown deviation g_α from a

Maxwellian still assuming quasi-neutrality. By inverting the equation for g_α , one is finally able to retrieve the kinetic equilibrium.

In many situations, for instance, the dayside magnetosphere, the Harris-like equilibrium (or any equilibrium obtained with the same mathematics) cannot be used because of the perfect symmetry of the Harris equilibrium with respect to the “neutral line” corresponding to the magnetic field reversal. Many attempts have been made in the past to overcome this problem in order to obtain a non-symmetric Vlasov equilibrium first of all addressing specifically the magnetosphere configuration (see [27] and refs. therein). Recently a *semi-analytic* kinetic model for a tangential current sheet has been developed [28]. This model considers kinetic ions and fluid electrons. It allows for a non-symmetric density with an antisymmetric magnetic field and a null electric field and valid even in the absence of a guide field. On the other hand, the electrons are assumed as fluid. As a consequence, only the ion d.f. is calculated. By means of a fully kinetic PIC approach, it has been shown that this kind of equilibrium remains stable even for an electron d.f. that does not correspond initially to a kinetic equilibrium [29]. More recently an exact 1D Vlasov analytical solution for an asymmetric Harris sheet has been proposed. The solution is obtained as a sum of four exponential functions of the constants of motion. The corresponding distribution function is given by the sum of four shifted Maxwellians [30].

A different approach to avoid the complications of starting with a Vlasov equilibrium is to take advantage from a kinetic equilibrium configuration generated, for example, by a primary instability then unstably to a secondary instability. This is the case, for example, of fast electron magnetic reconnection [31] developing on the shoulder of a sheared magnetic field generated by the development of the current filamentation (or Weibel) instability [32] (see Sect. 5.5.2 for details on this instability).

5.5 Basic Applications: Two Textbook Example

5.5.1 Electrostatic Limit: Landau Damping and Nonlinear Evolution

The electrostatic response of a plasma is a fundamental part of plasma dynamics which often do play a key role also in the more general electromagnetic context. One for all we cite the Landau damping process, a fundamental wave-particle interaction mechanism well known in the case of Langmuir waves, but of far-reaching consequences. For example, it turns out to be a key ingredient to calculate the correct dispersion relation of kinetic Alfvén waves (KAW) or of the whistler waves in the kinetic regime (both waves are at the basis of the most popular turbulence plasma models). Another fundamental electrostatic process, still not limited to the electrostatic dynamics, is particle trapping which can be seen as an alternative nonlinear saturation mechanism of instabilities (in particular streaming instabilities). Electrostatic dynamics is typically studied numerically by coupling

the Vlasov equation with the Poisson equation often assuming the ions just as a fixed neutralizing background of density n_0 because of the fast timescale of the electron dynamics. Limiting to the 1D-1V case and assuming periodic boundary conditions for the sake of simplicity, the system of equations reads:

$$\frac{\partial f}{\partial t} + v \frac{\partial f}{\partial x} + E \frac{\partial f}{\partial v} = 0 \quad (5.29)$$

$$\nabla^2 \phi = (n - n_0) \quad (5.30)$$

$$n = \int f dv; \quad E = -\nabla \phi$$

Here we have normalized all quantities using electron characteristic quantities, the plasma frequency ω_{pe} , the Debye length λ_D , and a characteristic electric field $E = mv_{th,e}\omega_{pe}/e$ where $v_{th,e}$ is the thermal velocity of the electrons. To solve numerically the above equations, only the d.f. needs to be initialized, $f(x, v, t = 0)$ depending on the problem to be studied. The electric field is given by the Poisson equation once $n(t = 0)$ has been calculated by the integral of the d.f., i.e., the zero moment of f . In the case of open boundary conditions instead, the Vlasov equation, Eq. (5.29), must be coupled to the Ampere equation, Eq. (5.5). This is an evolving equation for the electric field (in normalized units)

$$\frac{\partial E}{\partial t} = -j \quad (5.31)$$

where j is the electron current calculated as the first moment of the d.f. The Ampere equation must be used when dealing with open boundary conditions because using the Poisson equation one would instantaneously propagate any (unavoidable) numerical error at the boundary; see, e.g., [33]. However in this case the initial electric field $E(x, t = 0)$ must be properly set according the initialization of the d.f. in order to satisfy the Poisson equation at $t = 0$. Starting from the electron characteristic time of order of one in dimensionless units, $\tau_e = \omega_{pe}^{-1}$, the other two basic characteristic times characterizing the electrostatic regime are the Landau time, $\tau_L = \gamma_L^{-1}$, where γ_L is the Landau damping rate, and the particle bouncing time, $\tau_B = \omega_B^{-1}$ with $\omega_B = (kE_k)^{1/2}$ (still in dimensionless units). Here k and E_k are the wave number and the wave amplitude of a given fluctuation (typically a Langmuir wave). Analytically, one must solve the linearized Vlasov Poisson system³ which, in the Fourier space, corresponds to find the zero in the complex

³ To linearize the Vlasov Poisson system, one typically claims that $\delta f \ll f_0$ while what is actually assumed is $\partial_v \delta f \ll \partial_v f_0$. Since δf evolves rapidly generating steeper and steeper velocity gradient in phase space while the velocity gradient of the initial d.f. is fixed, after some time this assumption is no longer valid even if $\delta f \ll f_0$ is still satisfied as for a plasma wave.

plane of the dielectric function

$$D(k, \omega) = 1 + \frac{\omega_{pe}^2}{k} \int_{-\infty}^{\infty} \frac{\partial f_0 / \partial v}{\omega - kv} dv \quad (5.32)$$

where $\omega = \omega_r + i\gamma$ is the complex frequency and $\omega = kv$ corresponds to the poles. Analytically, the dispersion relation $D(k, \omega) = 0$ can be solved only in the limit $\omega \simeq \text{Re}(\omega)$, i.e., for $\omega_r \gg \gamma$ corresponding to the weakly damped regime:

$$D(\omega) \simeq D_r(\omega_r) + i\gamma \left(\frac{\partial D_r}{\partial \omega} \right)_{\omega=\omega_r} + i \left\{ D_i(\omega_r) + i\gamma \left(\frac{\partial D_i}{\partial \omega} \right)_{\omega=\omega_r} \right\} = 0$$

Since we assumed $\gamma \ll \omega_r$, we get

$$D_r(\omega_r) \simeq 0$$

where $\gamma \partial D_i / \partial \omega$ has been neglected because both γ and D_i are assumed small with respect to D_r (the velocity gradients are “gentle” in the vicinity of the phase velocity). The imaginary part gives:

$$\gamma = -\frac{D_i(\omega_r)}{\partial D_r(\omega)/\partial \omega}$$

Finally, by defining \mathcal{P} as the Cauchy principal value, we obtain:

$$D_r(k, \omega) = 1 + \frac{\omega_{pe}^2}{k} \mathcal{P} \int \frac{\partial f_0 / \partial v}{\omega - kv} dv; \quad D_i(k, \omega) = -\pi \frac{\omega_{pe}^2}{k} \frac{\partial f_0}{\partial v}$$

We take an initial 1D-1V Maxwellian velocity distribution

$$f_0 = (2\pi v_{th,e})^{-1/2} \exp[-v^2/2\pi v_{th,e}^2] \quad (5.33)$$

and assume (which could be questionable) that the phase velocity is much larger than the electron thermal velocity. Then, the integral can be evaluated explicitly, and one gets the following plasma wave dispersion relation (see [3] for the full theory):

$$\omega_r \simeq \omega_{pe} \left[1 + \frac{3}{2} (k\lambda_D)^2 \right]; \quad \gamma \equiv \gamma_L = -\sqrt{\frac{\pi}{8}} \frac{\omega_{pe}}{(k\lambda_D)^3} e^{-\left(\frac{3}{2} + \frac{1}{2k^2\lambda_D^2}\right)} \quad (5.34)$$

The plasma wave dispersion relation, Eq. (5.34), has been obtained in the limit $c_{ph} \gg v_{th,e}$ (where c_{ph} is the phase velocity) which corresponds to the limit $k\lambda_D \ll 1$. In this limit the timescale of the oscillation is much shorter than the damping time, $\tau_e \simeq 1/\omega_{pe} \ll \tau_L = 1/\gamma_L$ thus satisfying *a posteriori* the

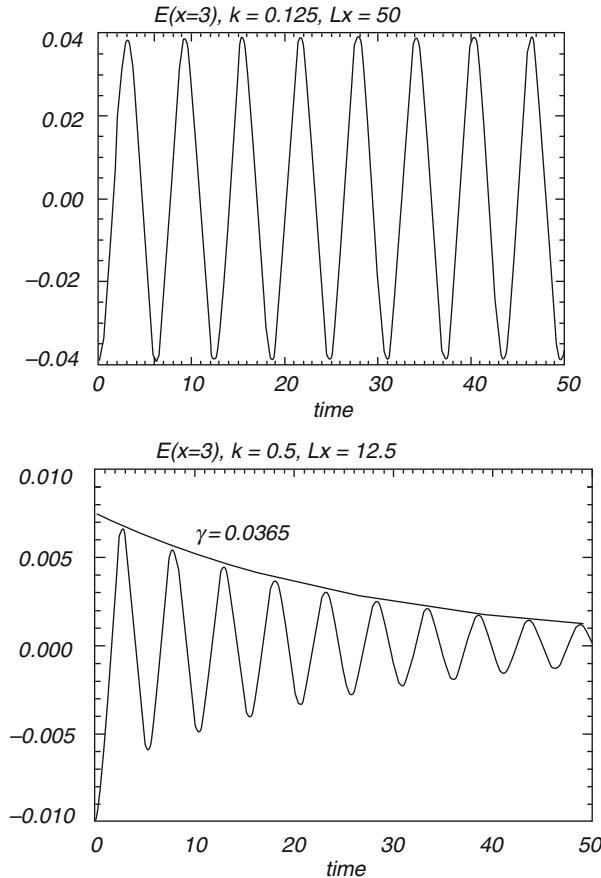


Fig. 5.3 The electric field oscillation by a plasma wave vs. time. Upper frame: undamped plasma wave, $k\lambda_D = 0.125$. Bottom frame: Damped plasma wave, $k\lambda_D = 0.6$

starting hypothesis of “weak” damping. Landau damping is a linear mechanism where electrons traveling at nearly the same speed of the wave can be accelerated or decelerated efficiently. Since the mechanism is collisionless, the energy of the electrons is not really damped but transferred and stored into small-scale oscillations of the d.f. In Fig. 5.3 we show an example of a 1D-1V simulation of a plasma wave made by taking a Maxwellian initial d.f., Eq. (5.33), supporting a sinusoidal fluctuation of wave number normalized to the Debye length much smaller than one, $k\lambda_D = 0.125$, left frame, or order one, $k\lambda_D = 0.6$, right frame. We see in the first case that the wave is undamped, as expected, while a weak damping is observed in the second case with a damping rate in excellent agreement with the theoretical estimation. For smaller wavelengths closer to the Debye length, the damping rate increases, while the analytical estimation (that assumes $c_\phi \gg v_{th,e}$) of the damping rate becomes less and less accurate. Particles with a kinetic energy less than the

potential energy of the wave will bounce back and forth into the potential well if the wave is damped more slowly than the bouncing time. This bouncing (or trapping) effect can be estimated roughly as follows. Let us assume a plasma wave given by

$$E(x, t) = \epsilon \sin(kx - \omega t); \quad \phi(x, t) = (\epsilon/k) \cos(kx - \omega t)$$

damped on a characteristic time $t_L = 1/\gamma_L$. An electron moving at nearly c_{ph} sees, in the wave reference frame, a quasi-stationary electrostatic field. As a result, the motion equation for the electron is given by

$$m \frac{d^2 x}{dt^2} = -e\epsilon \sin(kx) \simeq -e\epsilon kx \quad \rightarrow \quad T = \frac{1}{2\pi} \sqrt{\frac{m}{e\epsilon}} = \frac{2\pi}{\omega_B}$$

where ω_B is the bouncing frequency and T the period (assuming the regime $kx \ll 1$, i.e., $k\lambda_D \ll 1$). It can be shown that, by integrating along the unperturbed trajectory, the bouncing (trapping) time $\tau_B \equiv (m/e\epsilon)^{1/2}$ corresponds to the characteristic time for which the velocity derivative of the d.f. perturbation, $\partial(\delta f)/\partial v$, becomes of the same order as the velocity derivative of the initial Maxwellian, $\partial f_0/\partial v$ [3]. This is consistent with the fact that once the particle starts to be trapped, its trajectory is completely changed with respect to its initial state of a free traveling particle; in other words it corresponds to the transition to a nonlinear regime. This process can be studied numerically with a great accuracy by Eulerian Vlasov code (see [34] and references therein) and summarized as follows. Once particles start to be trapped, the d.f. develops a plateau around the resonant velocity (c_{ph}) and Landau damping stops since $\partial f_0/\partial v$ becomes locally null. In this case the wave amplitude, initially damped for $t < \tau_B$, remains constant without being damped anymore. The resulting particle trapping corresponds to harmonic-like oscillations of the particles in phase space leading the d.f. to form vortex-like structures, a process known as *nonlinear Landau damping*. In Fig. 5.4 we show a typical phase space vortex-like structure forming as soon as the nonlinear Landau damping comes into play, a clear signature of particles oscillating in phase space. More precisely, the figure is a zoom in the (x, v) phase space of the shaded iso-contours of the d.f. around the resonant velocity. We see that the corresponding isolines of the d.f. are wrapped more and more but remain quite well separated even if some numerical dissipation on the grid-scale length is at play because the velocity gradient scale became of the order of the grid mesh size. For this reason, it appears on the large scale as a closed vortex with well-defined separatrix.⁴ The numerical diffusion at the grid-scale length is unavoidable (and it would have been physical in any case since the increasing of the velocity gradient proceeds at an exponential rate [35]) and corresponds to a violation of the

⁴ The separatrix is a curve in phase space separating two distinct dynamical behavior, for instance, trapped particles inside a phase space vortex and free streaming particles outside. In the case of a plasma wave in the wave reference frame, trapped and free-streaming particles are those with kinetic energy less or greater than the wave potential energy.

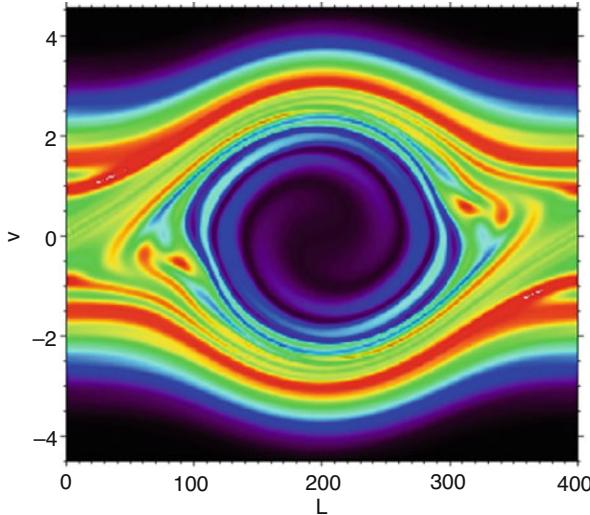


Fig. 5.4 Zoom of the shaded iso-contours of the distribution function in phase space around the resonant velocity resulting from a nonlinear Landau damping process

Vlasov equation. Such violation is very well visible on the Vlasov invariants, for example, the second- or third-order invariant; see Eq.(5.2) with $n = 2, 3$, which are very useful in Vlasov simulation as a check of the dissipative properties of the system. Indeed, these invariants will stay constant during the Landau damping process and the initial formation of the vortex-like structure in phase space. Then, as soon as the scale length of the fluctuations becomes comparable to the mesh size Δx , dissipating effects come into play. As a consequence, the invariants start rapidly to decrease till the vortex is formed by the breaking and reconnecting of the d.f. isolines (violation of the Vlasov equation). As soon as the vortex is well formed, the invariants stay again perfectly constant in time since phase space vortices correspond to Bernstein-Greene-Kruskal (BGK) wave solutions [36]. These structures are metastable Vlasov equilibria which is the reason why the invariants do not vary anymore. Finally, we note that phase space vortices driven by particle trapping is a general marker of a saturation process of electrostatic instabilities such as the bump on tail or the two stream instability, but at play also in the electromagnetic case as for the current filamentation instability (often known as the Weibel instability; see next section). Phase space vortex-like structures have been detected in the solar wind but also in the laboratory mainly in the form of bipolar electric signals and have been proposed as a possible collisionless “dissipative” mechanism.

5.5.2 Electromagnetic Case: The Current Filamentation (Weibel) Instability

The current filamentation instability (hereafter CFI) often referred as the Weibel instability (hereafter WI; see below) is an electromagnetic instability that generates a “quasi-static” magnetic field in the presence of two counter propagating electron populations (beams). This instability competes with the electrostatic two-stream instability and dominates when the beam velocity approaches the light speed. The growth rate increases for increasing values of the wave vector and saturates at around $k d_e \simeq 1$, corresponding to wavelengths of the order of the electron skin depth $d_e = c/\omega_{pe}$ (see [32] and references therein).

Typically, the initial counter-streaming beam configuration is achieved starting first with an electron beam generated by some acceleration process. Then, this beam is “immediately” compensated by a return current produced by the plasma to ensure quasi-neutrality. As a result, the initial total current is zero, $n_b \mathbf{v}_b + n_r \mathbf{v}_r = 0$ (where the index b and r refer to the *initial beam* and to the *return current*). Mostly in the laser-plasma interaction community (but not only), the CFI has been dubbed as the WI since the physical process underlying the instability development and the resulting magnetic field is very similar. Nevertheless, we recall that the WI is driven by an initial temperature anisotropy while the CFI by an initial electron momentum anisotropy. The development of the CFI is particularly relevant in the context of laser-plasma interaction where it was observed and studied more than 20 years ago, for instance, in the wake of an ultra-intense, ultra-short laser pulse propagating in an underdense plasma [37]. It became later, together with the WI, very popular also in the astrophysical context to explain the magnetic field structure observed in collisionless shocks [38, 39] or proposed as the mechanism capable of generating the initial magnetic field to be then amplified at cosmological level [40, 41].

The magnetic field generation process driven by the counter-propagating beams can be understood as follows. Let’s superpose to the initial currents directed along the y -axis of a Cartesian reference frame, an infinitesimal perpendicular perturbation with wave vector along the x -axis (see Fig. 5.5). The two currents are now infinitesimally displaced one with respect to the other. As a result, the initial perturbation is amplified since two opposite currents tend to repeal each other leading to the development of the instability feed by the energy initially stored in the two beams. The progressive separation of the currents corresponds to the growth of a dipolar magnetic field along the z -axis perpendicular to the initial beams and to the wave vector of the initial perturbation. It is worth noticing that the linear stage of the CFI can be in principle described in the framework of a fluid approach (see, e.g., [42] and references therein).

The kinetic study of the basic physics of the CFI and/or the WI can be done using an Eulerian Vlasov approach discussed in Sect. 5.2 limited to a reduced 1D-2V phase space. For beam velocities not exceeding $\sim 0.3c$ (where c is the speed of the light), relativistic effects can be neglected. The spatial direction is aligned with the most unstable wave vectors, and the 2D phase space is the minimum needed for

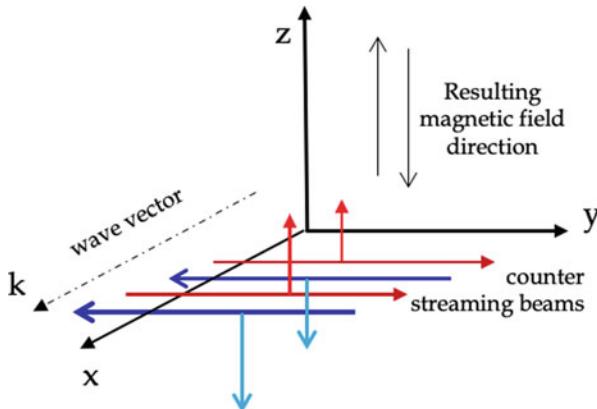


Fig. 5.5 Schematic representation of the current filamentation instability

allowing the gyration of the particles around the generated magnetic field. In this configuration the two-stream instability, which develops along the same direction of the beams, is therefore excluded a priori. Furthermore, since the instability develops on the electron timescale, one can assume the ions as a fixed neutralizing background. We assume periodic boundary conditions. By defining $f_e(x, v_x, v_y, t)$ the electron velocity distribution function (hereafter EVDF), the full set of equation reads:

$$\frac{\partial f_e}{\partial t} + \mathbf{v}_e \cdot \frac{\partial f_e}{\partial \mathbf{x}} + (\mathbf{E} + \mathbf{v}_e \times \mathbf{B}) \cdot \frac{\partial f_e}{\partial \mathbf{v}_e} = 0$$

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E}; \quad \frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{B} - \mathbf{j} \quad (5.35)$$

In the above equations, we have normalized all quantities using the electron mass m_e , a characteristic particle density \bar{n} , the speed of light c , the electron plasma frequency $\omega_{pe} = (4\pi\bar{n}e^2/m_e)^{1/2}$, and the characteristic electric and magnetic fields $\bar{E} = \bar{B} = m_e c \omega_{pe}/e$. The initial EVDF carrying the two initial counter-propagating currents is given by:

$$f_e(v_x, v_y, t = 0) = \frac{1}{\pi u_{th}^2} e^{-(v_x/u_{th})^2} \left[\mu e^{-[(v_y+u_{0,1})/u_{th}]^2} + (1-\mu) e^{-[(v_y+u_{0,2})/u_{th}]^2} \right]$$

where $u_{0,1/2}$ is the beam velocity of the first/second beam, $1/2$, respectively. Furthermore, $\mu \in (0, 1)$ represents the beam density asymmetry (symmetric for $\mu = 0.5$), and u_{th} is the thermal velocity which is taken equal for the two beams for the sake of simplicity. We recall that the total electron density is normalized to one. Since the net initial current must be zero, we have $\mu u_{0,1} + (1-\mu) u_{0,2} = 0$. To initialize the numerical simulation in the simplest way, one must set to zero

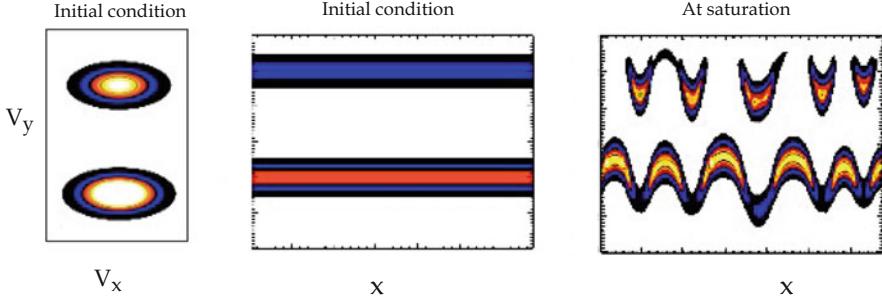


Fig. 5.6 Schematic representation of the current filamentation instability development in phase space. Left and middle frame: the EVDF at the initial time in the (v_x, v_y) and (x, v_y) phase space. Right frame: the EVDF at saturation in the (v_x, v_y) phase space. Figures from [43] are reprinted with permission from Nova Science Publishers, Inc.

the initial electric field and adding a very small sinusoidal magnetic perturbation along the z direction of the form $\delta b_z = \epsilon \sin(kx)$, with $\epsilon \ll 1$. Once the perturbation amplitude grows, the beams start to be deflected more and more in the perpendicular direction, i.e., along v_x . The instability saturates kinetically as soon as the gyroradius becomes of the order of d_e since the initial momentum anisotropy is lost.

In Fig. 5.6 we show a simple sketch of the EVDF evolution in the case of the CFI development. In the first and second frame, we plot the two initial beams in the (v_x, v_y) and (x, v_y) phase space. The beams have a mean velocity $u_{0,1} = -0.1$ and $u_{0,2} = 0.2$ normalized to the light speed. The space interval is $L = 12\pi$ in d_e units. The middle picture shows that the beams are homogeneous in space. After the development of the instability, the most unstable mode $k \simeq 2\pi$ generates magnetic field peaks around which the “particles” (i.e., the EVDF) start to spread leading at saturation (right frame) to a more and more isotropic configuration of the beams in their respective reference frame. A complete analysis of the kinetic saturation of the CFI can be found in [32].

5.6 Advanced Applications: Recent Progress in Space Plasma Turbulence via Eulerian Vlasov Simulations

Plasma turbulence in weakly collisional space and astrophysical environments, as well as its interplay with magnetic reconnection, represents today a formidable scientific challenge in plasma research and in computational physics. A very large number of theoretical and numerical efforts have been made in the last decades in order to understand these physical processes.

5.6.1 Plasma Turbulence from In situ Measurements in the Solar Wind and in the Earth’s Magnetosheath: A Brief Overview

Alongside theory and simulations, a fundamental role has been played by increasingly sophisticated and accurate space missions that have been employed to this end, such as the Cluster [44] and Magnetospheric Multiscale (MMS) [45] missions. The near-Earth environment and the solar wind represent the best laboratory where plasma turbulence and magnetic reconnection can be studied in great details via in situ measurements of particle distributions and electromagnetic fields [7, 46–53]. Such observations represent a sort of polar star for many numerical and theoretical studies of space plasmas. Spacecraft data indeed provide important observational constraints, for instance, on the typical spectral energy distribution of the turbulent fluctuations and on the turbulent energization and heating of the plasma components. In particular, the electromagnetic fields exhibit a cascade of energy from “large” fluid (MHD) scales down to the “small” ion and electron kinetic scales, where such fluctuation energy is eventually “dissipated,” i.e., converted into plasma heat. In fact, while measurements of such fluctuations at MHD scales show an anisotropic (with respect to the magnetic field direction [54]) cascade of mainly *critically balanced* Alfvénic fluctuations⁵ [56–60], with the typical $-5/3$ slope in wavenumbers perpendicular to \mathbf{B} , the nature of the cascade changes as soon as the plasma microscales are encountered [55, 61, 62]. A first spectral break emerges at ion-kinetic scales, after which the magnetic field spectrum steepens, exhibiting a typical spectral slope around -2.8 , and the spectrum of the electric-field fluctuations simultaneously flattens [49, 51, 52, 63–65]. A second break, with an associated further steepening of the fluctuation spectra, is finally observed at the electron scales [51, 52]. Understanding the fundamental mechanisms underlying the observed energy transfer through scales, as well as the damping mechanisms of such turbulent fluctuations, is a fundamental step to take if we want to understand phenomena such as the observed non-adiabatic heating of the solar wind [66–68] and the energy partition between species [69–71]. In contrast with the case of neutral fluids, in a weakly collisional plasma, the turbulent energy transfer may involve many different processes such as fluid and kinetic instabilities that compete with the standard “hydrodynamic-like” nonlinearities (i.e., local “wave-wave” interaction). In the context of wave-like models of the kinetic-scale turbulent cascade in a plasma, the most invoked ones have been those based on kinetic Alfvén waves (KAWs) and/or whistler waves [55, 61, 72, 73]. However, the fact that the spectral slopes observed in the solar wind are systematically steeper than those predicted by such phenomenological theories (e.g., $-7/3$ for the magnetic field spectrum) requires us to include additional effects, such as dissipation and/or intermittency [69, 74, 75]. Within this scenario, it has also been recently proposed that magnetic reconnection

⁵ This means that the fluctuations reach a balance between their linear Alfvén-wave propagation timescale along B , i.e., $\tau_A \sim (k_{\parallel} v_A)^{-1}$, and their nonlinear interaction timescale (“cascade time”) in the field-perpendicular direction, i.e., $\tau_{nl} \sim (k_{\perp} \delta u_{\perp,k})^{-1}$ [54, 55].

could indeed play a fundamental role in the cascade model by competing with and/or completely replacing the standard wave-like mechanism of nonlinear energy transfer [76–79]. In particular, recent hybrid-Vlasov and hybrid-PIC simulations have shown that the continuous formation of current sheets and their disruption via fast (“collisionless”) reconnection likely supports a *non-local energy transfer* from fluid to sub-ion scales, possibly feeding also a simultaneous *inverse cascade* back to MHD scales [77]. The idea behind this mechanism is that the large-scale turbulent motions continuously generate elongated current sheets whose rapid shrinking into thin-and-elongated structures directly provides fluctuations at the ion kinetic scales (viz., the ion Larmor radius and/or the ion skin depth depending on the plasma beta [80]). Then, their disruption via reconnection very efficiently populates the turbulent spectrum at sub-ion scales, either via local (in space) injection of wave-like fluctuations and by growing coherent structures (magnetic islands and flux tubes). These structures may also interact and eventually merge, thus feeding a generation of growing-in-size structures (viz., an inverse cascade). This reconnection-mediated process can be viewed as an entirely new paradigm in turbulence theory; a (to-be-further-confirmed) hint of its signatures has been recently spotted using solar-wind data [81].

We also mention that, on top of the above processes, the turbulent dynamics naturally leads to pressure anisotropies that in turn can be an efficient source of instabilities, for instance, the mirror and firehose instability (e.g., [66–68, 71, 82–87]). Therefore, understanding the competition and interaction between the nonlinear turbulent cascade and fluid/kinetic instabilities is today the main challenge in developing a theory for magnetized turbulence in collisionless plasmas.

5.6.2 Plasma Turbulence from Kinetic Simulations

A rigorous approach to this intrinsically nonlinear and multi-scale problem would require the so-called fully kinetic model. Such approach consists of solving the full Vlasov–Maxwell system of equations, Eqs. (5.1)–(5.4)–(5.5), in order to self-consistently determine the evolution of all the species’ particle distribution function in phase space and of the electromagnetic fields in position space. However, from a computational point of view, a fully kinetic description would inevitably face the curse of an extremely high dimensionality, as well as the necessity to resolve a wide range of space and temporal scales which for realistic parameters usually span many orders of magnitude. Therefore, a common approach to the problem is to adopt the so-called reduced-kinetic models that represent limits of the fully kinetic case that may hold under certain circumstances. Important progress in our understanding of kinetic plasma turbulence has been indeed achieved by means of the numerical solution of reduced versions of the full Vlasov–Maxwell equations. In general, we remind to the reader that the practical approach to the solution of the (full or reduced) Vlasov equation relies on two different numerical methods introduced in the previous sections: (i) a Lagrangian approach, whose implementations are usually referred to as PIC codes (see Chap. 6 of this book), or (ii) an Eulerian method

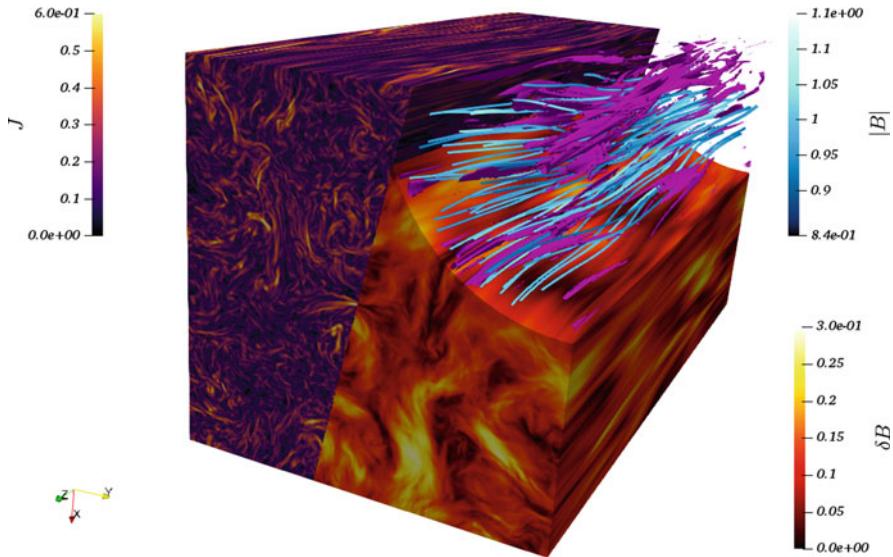


Fig. 5.7 Three-dimensional rendering of current density, J , and magnetic field fluctuations, δB . Magnetic field lines with color-coded magnetic field strength, $|B|$, are also shown, along with embedded current density iso-surfaces (purple). The dataset used to produce this image is from a $\beta = 1$ hybrid-Vlasov–Maxwell (HVM) simulation of plasma turbulence (see Refs. [88, 89]). The absence of noise in Vlasov algorithms is clearly appreciable for turbulence studies, especially in the current density, $\mathbf{J} = \nabla \times \mathbf{B}$. However, it clearly comes at a considerable cost in terms of computational resources

(this Chapter) that is implemented in the so-called “Vlasov” codes (Fig. 5.7). In particular, recent efforts have been put into comparisons between these different numerical approaches, as well as between different reduced models and a fully kinetic approach [90–92]. We stress here that choosing between the Lagrangian or the Eulerian approach is just a matter of choosing between a different numerical method that solves the Vlasov equation by advancing the distribution function along the characteristics in phase space (while no actual particle trajectories are integrated at all). As a matter of fact, each of these two approaches is better suited for investigating different processes and/or regimes, and so they should be considered as complementary rather than competitive.

5.6.2.1 Decreasing the Dimensionality of the Problem: Plasma Turbulence via Reduced-Kinetic Eulerian Simulations

Among reduced-kinetic models and their application to investigate the turbulent dynamics in space plasmas, we mention that a fairly recent approach is represented by the so-called gyrokinetic (GK) model (see [55, 69, 71] and references therein). Such model is a reduced-kinetic framework that has been typically employed by the magnetic-confinement-fusion community to study the low-frequency, small-amplitude turbulent fluctuations in (strongly) magnetized plasmas (see, e.g., [93]

and references therein). In the GK model, the gyration of the charged particles around the magnetic field represents one of the fastest (“ignorable”) processes, and thus, within the hypothesis known as “gyrokinetic ordering” it can be removed from the plasma description [94, 95]. That ordering in practice excludes from the plasma description the high-frequency modes such as the fast-magnetosonic and the whistler waves, the wave-particle interaction related to the cyclotron resonances, and finite-amplitude effects at the Larmor scales, for instance, the so-called stochastic heating. The underlying idea of the gyrokinetic approach to space plasma turbulence is essentially based on the spectral anisotropy characterizing a critically balanced cascade of Alfvénic fluctuations at MHD scales [54]. Because of such anisotropy, the turbulent energy is mainly cascading toward small perpendicular (to \mathbf{B}) scales. In that case, it is believed that when the cascade reaches the (ion) kinetic scales, the associated turbulent dynamics would be dominated by very oblique, low-frequency, small-amplitude KAW-like fluctuations. One of the advantage of the GK description is that, thanks to the great computational simplification associated to its approximation, it has been possible to perform 3D Eulerian gyrokinetic simulations of weakly collisional plasma turbulence covering a relatively wide spectral range and yet adopting a realistic ion-to-electron mass ratio [96–99]. Within the limits of validity of the GK model, a certain amount of spectral properties of the fluctuations obtained from those simulations turn out to be qualitatively in agreement with observations. For further details on the gyrokinetic model, its numerical implementation and applications, see Chap. 4 of this book.

Still along the line of reduced-kinetic models, Eulerian simulations of the Vlasov–Maxwell system in the hybrid approximation (see Sect. 5.3) have been widely employed in the recent years to study sub-ion-range turbulence and the transition across the ion cyclotron frequency. Most of these works have been focusing on freely decaying fluctuations in a proton-electron plasma and within a reduced 2D-3V phase space [100–106]. Those pioneering studies investigated, among other things, the fundamental role of current sheets in kinetic plasma turbulence. Current sheets are the structures that are mostly related to energy injection at sub-ion scale, development of strong anisotropies, and non-Maxwellian features in the distribution function. Even if Vlasov studies are already computationally demanding, the further advances in computational capabilities of recent years have made possible to extend those works to more complex cases. Still in 2D, it was possible to add a second ion (kinetic) species made of α particles [107–109] and to show their differential heating and anisotropy production. Another aspect that has been investigated via two-dimensional hybrid-Vlasov simulations is the transition to kinetic-scale turbulence by large-amplitude kinetic-Alfvén waves (KAWs) in an inhomogeneous background [110]. Moreover, recent 2D-3V investigations included an external drive in order to achieve a quasi-steady turbulent state over which turbulence properties could be statistically averaged (see, e.g., Fig. 5.8). Such studies showed that the polarization of sub-ion-scale turbulent fluctuations is indeed changing with the plasma β parameter [111] and that such effect takes place (largely) regardless of their original properties at MHD scales [90]. Furthermore, those simulations reported the first evidence that initiated the change of paradigm in

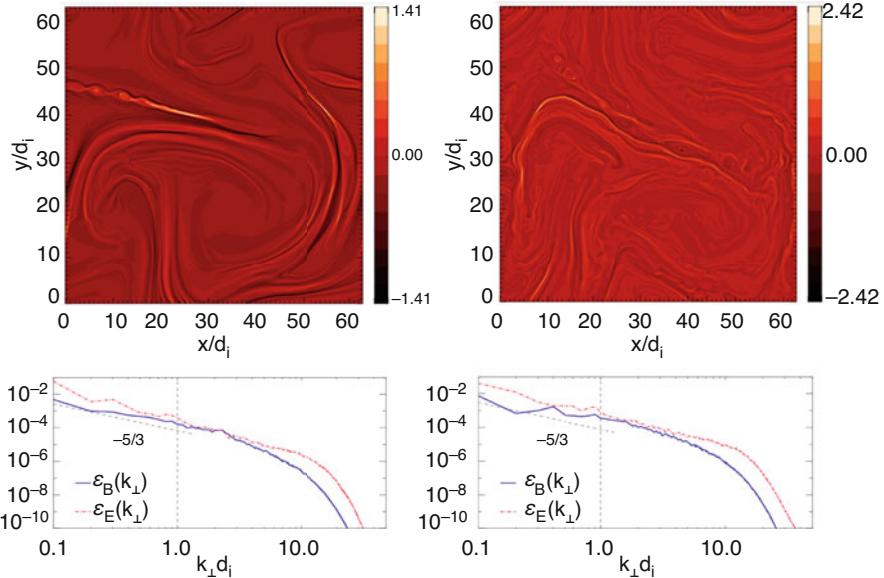


Fig. 5.8 Contour plots of the out-of-plane current density, J_z (top row), and one-dimensional electric and magnetic power spectra, $\mathcal{E}_E(k_\perp)$ and $\mathcal{E}_B(k_\perp)$ (bottom row), at two distinct times (left and right column, respectively) in a 2D3V hybrid-Vlasov simulation of (continuously driven) turbulence, spanning from MHD to sub-ion scales. This simulation highlighted a direct link between the timescale of fast magnetic reconnection and the nonlinear energy transfer timescale associated to kinetic-range turbulence (now known as *reconnection-mediated* turbulence). The figures are adapted from [76], reproduced with permissions under the CC BY license

turbulence theory referred to as “*reconnection-mediated (kinetic) turbulence*”, i.e., that magnetic-reconnection timescales may become the relevant nonlinear timescale that is responsible for the turbulent energy transfer at and below the ion kinetic scales [76, 77]. This was highlighted for the first time in [76] where the authors originally proposed the idea that “*the formation of coherent structures by fast reconnection processes can be the crucial mechanism that continue the nonlinear cascade across the ion break*” (Fig. 5.8). Based on such 2D-3V hybrid-Vlasov simulations of continuously driven turbulence, they conjectured the now-accepted idea that such regime would only require “*the occurrence of ‘fast enough’ magnetic reconnection events developing on a timescale shorter than (or comparable to) those of the wave mode interactions*” [76]. A more robust numerical proof was eventually provided in [77]. A typical signature of this reconnection-mediated regime in kinetic-range turbulence, for instance, is a magnetic-fluctuations’ spectrum whose spectral slope is between $-8/3$ and -3 [76–79] (significantly steeper than the $k_\perp^{-7/3}$ predicted by standard KAW turbulence [55]). The scale at which the magnetic spectrum

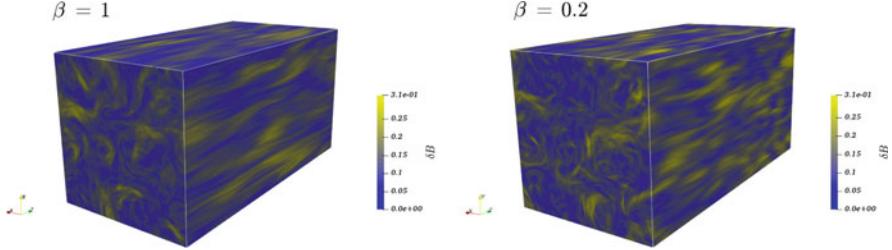


Fig. 5.9 3D contours of the magnetic field fluctuations, δB , from 3D-3V HVM simulations during developed turbulence for $\beta = 1$ (left) and $\beta = 0.2$ (right) regimes. See Refs. [88, 92] and references therein for further details about intermittency and spectral anisotropy of the turbulent fluctuations

breaks in the reconnection-mediated regime, called the “disruption scale”,⁶ also has a specific dependence on plasma parameters [78, 79, 81]. Although some pieces of this “puzzle” have been placed together, a sharper understanding of this regime will need to await further numerical, theoretical, and observational studies. In particular, the numerical approach would be computationally demanding. One should clearly identify and separate the disruption scale from the ion-kinetic scales, which requires to explore a wide range of plasma parameters and to possibly do so in 3D. Such a systematic and detailed investigation is currently out of computational reach, even in the context of hybrid-kinetic models.

Finally, only very recently it has become feasible to perform high-resolution 3D-3V hybrid-Vlasov simulations. After the first pioneering, lower-resolution, three-dimensional Vlasov-Eulerian simulations presented in [104], recent 3D studies spanning more than two decades of perpendicular (to \mathbf{B}_0) wavenumbers and including also electron-inertia effects were providing significant insights in sub-ion-range plasma turbulence [88, 89, 92]. Such studies confirmed that the change of the polarization exhibited by sub-ion-scale fluctuations is in fact changing with the plasma β [88] (cf. Fig. 5.9), as originally suggested via 2D simulations [90, 111], and that at $\beta \sim 1$ some of the spectral properties of these fluctuations (viz., slopes and anisotropy) are roughly consistent with a theory of intermittency corrected KAW turbulence presented in [74] (see [88] for the original work and [92] for more recent developments). This is, in a sense, also consistent with the long-suggested inherent relevance of current sheets (and their subsequent undergoing to magnetic reconnection) in kinetic plasma turbulence by 2D-3V Vlasov simulations [76, 77, 90, 101–109]. Moreover, via a mixed Fourier-Hermite representation

⁶ After the disruption of current (sheet) structures due to magnetic reconnection.

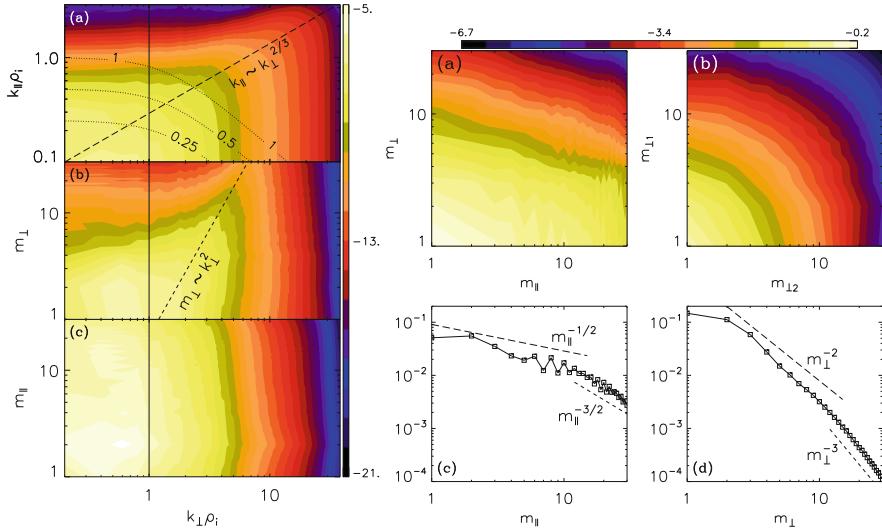


Fig. 5.10 Left panels: Mixed Fourier-Hermite spectrum (“*phase-space spectrum*”) of the fluctuating distribution function, $\log |\delta f_{k,m}|^2$. Right panels: Hermite spectrum of the fluctuating distribution function, $\log |\delta f_{m,l}|^2$, in a coordinate system aligned with the *local* magnetic field. See [89] and references therein for further details about the mixed Fourier-Hermite representation and about the scalings shown in the plots. Note that this is another point of strength of Eulerian algorithms: representing the distribution function f on a grid in phase space allows a clean and complete extraction of v -space information in each spatial point. Figures from [89] (©AAS. Reproduced with permission)

of the distribution function⁷ from these 3D-3V simulations, it has been possible to show for the first time that kinetic turbulence implies the existence of a turbulent cascade involving the entire six-dimensional phase space [89] (i.e., a 6D anisotropic, with respect to \mathbf{B} , cascade of ion-entropy fluctuations simultaneously taking place in position and in velocity space; see Fig. 5.10). Such “*phase-space cascades*” are indeed shown to be a process that inherently entwines its real and velocity space components [55, 89, 112] (cf. Fig. 5.10). We mention that similar results, previously obtained only within the gyrokinetic framework in the electrostatic limit and in a reduced 2D(2V) geometry [113, 114], have been independently shown

⁷ The Hermite transform of the distribution function (or, for instance, of its fluctuations, $\delta f = f - F_M$, with respect to a local Maxwellian, F_M) is a representation of f in terms of (orthogonal) Hermite polynomials of order m , $H_m(v)$, as basis functions in velocity space v . The Hermite representation in velocity space is analogous to the Fourier representation in real space x : higher Hermite numbers m describe finer structures of f in v space, the same way as higher Fourier modes k describe finer structures in real space. The Hermite representation is a suitable choice for representing a distribution function in velocity space because the first three Hermite numbers, $m = 0, 1, 2$, have a direct connection with the first three v -space moments of a Maxwellian distribution (*viz.*, number density n , mean flow u , and temperature T , respectively).

to exist also in 2D-3V hybrid-Vlasov simulations [115] and in 3D-2V hybrid-gyrokinetic ones [116]. More importantly, the Hermite decomposition applied to the ion distribution function obtained from in situ measurements in the Earth's turbulent magnetosheath by MMS was indeed able to provide the first experimental evidence that such velocity space cascade is actually taking place in turbulent plasmas at kinetic scales [117].

Finally, the recent advances in numerical algorithms development and their application to the Vlasov equation allowed to study fully kinetic 2D-3V turbulence within a Vlasov (Eulerian) approach [16]. Computational constraints are clearly still limiting the range and the separation of scales that can be investigated in terms of resolution achievable, MHD scales that can be included, and reduced ion-electron mass ratio that can be used. This problem still remains the curse of 3D-3V Vlasov simulations, even in the hybrid approximation. Nevertheless, those studies allowed to gain significant insights on kinetic-range turbulence, as well as to pose new questions (see discussion in [92]). To go beyond these achievements and to possibly answer those open questions, next-generation simulations will require a significant investment in computational resources, as well as the development of increasingly efficient numerical algorithms.

References

1. Cheng C., Knorr G., The integration of the Vlasov equation in configuration space. *J. Comp. Phys.* **22**, 330 (1976)
2. Gagne R., Shoucri M., A splitting scheme for the numerical solution of the Vlasov equation. *J. Comp. Phys.* **24**, 445 (1977)
3. N.A. Krall, A.W. Trivelpiece, *Principles of Plasma Physics* (San Francisco Press, 1986)
4. Oteo J.A., Ros J., The Magnus expansion for classical Hamiltonian systems. *J. Phys. A* **24**, 5751 (1991)
5. A. Mangeney, F. Califano, C. Cavazzoni, A numerical scheme for the integration of the Vlasov–Maxwell system of equations. *J. Comp. Phys.* **179**, 495 (2002)
6. F. Valentini, F. Califano, A. Mangeney, A Hybrid-Vlasov model based on the current advance method for the simulation of collisionless magnetized plasma. *J. Comp. Phys.* **225**, 753 (2007)
7. F. Sahraoui, M.L. Goldstein, G. Belmont, P. Canu, L. Rezeau, Three dimensional anisotropic k spectra of turbulence at sub-proton scales in the solarwind. *Phys. Rev. Lett.* **105**, 131101 (2010)
8. D. Winske, Hybrid simulation codes with application to shocks and upstream waves. *Space Sci. Rev.* **42**, 5 (1985)
9. Matthews A.P., Current advance method and cyclic leapfrog for 2D multispecies hybrid plasma simulations. *J. Comp. Phys.* **112**, 102 (1994)
10. N.V. Elkina, J. Büchner, A new conservative unsplit method for the solution of the Vlasov equation. *J. Comput. Phys.* **213**(2), 862–875 (2006)
11. M. Palmroth, U. Ganse, Y. Pfau-Kempf et al., Living Rev. Comput. Astrophys 4:1 (2018). <https://doi.org/10.1007/s41115-018-0003-2>
12. D. Pokhotelov, S. Von Alfthan, Y. Kempf, R. Vainio, H.E.J. Koskinen, M. Palmroth, Ion distributions upstream and downstream of the Earth's bow shock: first results from Vlasiator. *Ann. Geophys.* **31**, 2207–2212 (2013). <https://doi.org/10.5194/angeo-31-2207-2013>

13. S. Von Alfthan, D. Pokhotelov, Y. Kempf, S. Hoilijoki, I Onkonen, A. Sandroos, M. Palmroth, Vlasiator: first global hybrid-Vlasov simulations of Earth's foreshock and magnetosheath. *J. Atmosph. and Solar-Terrestrial Phys.* **120**, 24 (2014)
14. L. Palodhi, F. Califano, F. Pegoraro, Nonlinear kinetic development of the Weibel instability and the generation of electrostatic coherent structures. *Plasma Phys. Controlled Fusion* **51**, 125006 (2009)
15. L. Palodhi, F. Califano, F. Pegoraro, On the transition between the Weibel and the whistler instabilities. *Plasma Phys. Controlled Fusion* **52**, 095007 (2010)
16. J. Juno, A. Hakim, J. TenBargea, E. Shib, W. Dorland, Discontinuous Galerkin algorithms for fully kinetic plasmas. *J. Comp. Phys.* **353**, 110 (2018)
17. O. Pezzi, G. Cozzani, F. Califano, F. Valentini, M. Guarriasi, E. Camporeale, G. Brunetti, A. Retinò, P. Veltri, AViDA: a Vlasov-DArwin solver for plasma physics at electron scales. *J. Plasma Phys.* (in press)
18. H. Schmitz, R. Grauer, Darwin–Vlasov simulations of magnetised plasmas. *J. Comput. Phys.* **214**, 738 (2006)
19. E.G. Harris, *Il Nuovo Cimento* **23**, 115 (1962)
20. M.G. Green, *Phys. Fluids B* **5**, 1715 (1993)
21. J. Birn et al., Geospace environmental modeling (GEM) magnetic reconnection challenge. *J. Geophys. Res.* **106** (A3), 3715–3719 (2001)
22. J. Büchner, J.P. Kuska, Numerical simulation of three-dimensional reconnection due to the instability of collisionless current sheets. *Adv. Space Res.* **19**(12), 1817 (1997)
23. B. Coppi, J.W. Mark, L. Sugiyama, G. Bertin, *Ann. Phys.* **119**, 370 (1979)
24. J. Chen, P. Palmadesso, *Phys. Fluids* **27**, 1198 (1984)
25. N. Attico, F. Pegoraro, Periodic equilibria of the Vlasov–Maxwell system. *Phys. Plasmas* **6**, 767 (1993)
26. O. Allanson, T. Neukirch, S. Troscheit, F. Wilson, From one-dimensional fields to Vlasov equilibria: theory and application of Hermite polynomials. *J. Plasma Phys.* **82** (2016)
27. M. Roth, J. de Keyser, M.M. Kuznetsova, Vlasov Theory of tangential discontinuities. *Space Sci. Rev.* **76**, 251 (1996)
28. G. Belmont, N. Aunai, R. Smets, *Phys. Plasmas* **19**, 022108 (2012)
29. Dargent, J., Aunai, N., Belmont, G., et al., *J. Plasma Phys.* **82** (2016)
30. O. Allanson, F. Wilson, T. Neukirch, Y.-H. Liu, J.D.B. Hodgson, Exact Vlasov–Maxwell equilibria for asymmetric current sheets. *Geophys. Res. Lett.* **44**(17), 8685–8695 (2017)
31. F. Califano, N. Attico, F. Pegoraro, G. Bertin, S. Bulanov, Fast formation of magnetic islands in a plasma in the presence of counter streaming electrons. *Phys. Rev. Lett.* **86**, 5293 (2001)
32. F. Califano, F. Pegoraro, S. Bulanov, A. Mangeney, Kinetic saturation of the Weibel instability in a collisionless plasmas. *Phys. Rev. E* **57**, 7048 (1998)
33. F. Califano, M. Lontano, Electron hole generation and propagation in a inhomogeneous collisionless plasma. *Phys. Rev. Lett.* **95**, 245002 (2005)
34. M. Brunetti, F. Califano, F. Pegoraro, Asymptotic evolution of nonlinear Landau damping. *Phys. Rev. E* **62**, 4109 (2000)
35. L. Galeotti, F. Califano, Asymptotic evolution of weakly collisional Vlasov–Poisson plasmas. *Phys. Rev. Lett.* **95**, 015002 (2005)
36. M. Bernstein, J.M. Greene, M.D. Kruskal, *Phys. Rev.* **108**, 546 (1957)
37. T. Liseikina, F. Califano, V. Vshivkov, F. Pegoraro, S. Bulanov, Small scale electron density and magnetic field structures in the wake of a ultra intense laser pulse. *Phys. Rev. E* **60**, 599 (1999)
38. A. Spitkovsky, *Astrophys. J.* **673**, L39 (2008)
39. C. Huntington, F. Fiuzza, J. Ross, et al., Observation of magnetic field generation via the Weibel instability in interpenetrating plasma flows. *Nat. Phys.* **11**, 173–176 (2015)
40. F. Califano, T. Cecchi, C. Chiuderi, Nonlinear kinetic regime of the Weibel instability in an electron-ion plasma. *Phys. Plasmas* **9**(2), 451–457 (2002)
41. M. Lazar, R. Schlickeiser, R. Wielebinski, S. Poedt, *Astrophys. J.* **693**, 1133 (2009)

42. F. Califano, F. Pegoraro, S. Bulanov, Spatial structure and time evolution of the Weibel instability in collisionless inhomogeneous plasmas. *Phys. Rev. E* **56**(1), 963 (1997)
43. M. Shoucri (ed.), *Eulerian Codes for the Numerical Solution of the Kinetic Equations of Plasmas* (Nova Science Publishers, New York, 2011)
44. C.P. Escoubet, R. Schmidt, M.L. Goldstein, Cluster—science and mission overview. *Space Sci. Rev.* **79**, 11–32 (1997)
45. J.L. Burch, T.E. Moore, R.B. Torbert, B.L. Giles, Magnetospheric multiscale overview and science objectives. *Space Sci. Rev.* **199**, 5–21 (2016)
46. R. Bruno, V. Carbone, The solar wind as a turbulence laboratory. *Living Rev. Solar Phys.* **10**, 2 (2013)
47. M. Øieroset, T.D. Phan, M. Fujimoto, R.P. Lin, R.P. Lepping, In situ detection of collisionless reconnection in the Earth’s magnetotail. *Nature* **412**, 414 (2001)
48. A. Vaivads, Y. Khotyaintsev, M. André, A. Retinó, S.C. Buchert, B.N. Rogers, P. Décréau, G. Paschmann, T.D. Phan, Structure of the magnetic reconnection diffusion region from four-spacecraft observations. *Phys. Rev. Lett.* **93**, 105001 (2004)
49. S.D. Bale, P.J. Kellogg, F.S. Mozer, T.S. Horbury, H. Reme, Measurement of the electric fluctuation spectrum of magnetohydrodynamic turbulence. *Phys. Rev. Lett.* **94**, 215002 (2005)
50. D. Sundkvist., A. Retinó, A. Vaivads., S.D. Bale, Dissipation in turbulent plasma due to reconnection in thin current sheets. *Phys. Rev. Lett.* **99**, 025004 (2007)
51. F. Sahraoui, M.L. Goldstein, P. Robert, Y.V. Khotyaintsev, Evidence of a cascade and dissipation of solar-wind turbulence at the electron gyroscale. *Phys. Rev. Lett.* **102**, 231102 (2009)
52. O. Alexandrova, C.H.K. Chen, L. Sorriso-Valvo, T.S. Horbury, S.D. Blae, Solar wind turbulence and the role of ion instabilities. *Space Sci. Rev.* **178**, 101 (2013)
53. J.L. Burch et al., Electron-scale measurements of magnetic reconnection in space. *Science* **10**, 1126 (2016)
54. P. Goldreich, S. Sridhar, Toward a theory of interstellar turbulence. 2: strong alfvenic turbulence. *Astrophys. J.* **438**, 763–775 (1995)
55. A.A. Schekochihin, S.C. Cowley, W. Dorland, G.W. Hammett, G.G. Howes, E. Quataert, T. Tatsuno, Astrophysical gyrokinetics: kinetic and fluid turbulent cascades in magnetized weakly collisional plasmas. *Astrophys. J. Suppl. Series* **182**, 310 (2009)
56. M.L. Goldstein, D.A. Roberts, W.H. Matthaeus, Magnetohydrodynamic turbulence in the solar wind. *Annu. Rev. Astron. Astrophys.* **33**, 283 (1995)
57. T.S. Horbury, M. Forman, S. Oughton, Anisotropic scaling of magnetohydrodynamic turbulence. *Phys. Rev. Lett.* **101**, 175005 (2008)
58. J.J. Podesta, Dependence of solar-wind power spectra on the direction of the local mean magnetic field. *Astrophys. J.* **698**, 986 (2009)
59. R.T. Wicks, T.S. Horbury, C.H.K. Chen, A.A. Schekochihin, Power and spectral index anisotropy of the entire inertial range of turbulence in the fast solar wind. *Mon. Not. R. Astron. Soc.* **407**, L31 (2010)
60. C.H.K. Chen, A. Mallet, T.A. Yousef, A.A. Schekochihin, T.S. Horbury, Anisotropy of Alfvénic turbulence in the solar wind and numerical simulations. *Mon. Not. R. Astron. Soc.* **415**, 3219 (2011)
61. S. Boldyrev., K. Horaites, Q. Xia, J. Perez, Toward a theory of astrophysical plasma turbulence at subproton scales. *Astrophys. J.* **777**, 41 (2013)
62. C.H.K. Chen, Recent progress in astrophysical plasma turbulence from solar wind observations. *J. Plasma Phys.* **82**, 535820602 (2016)
63. C.H.K. Chen, T.S. Horbury, A.A. Schekochihin, R.T. Wicks, O. Alexandrova, J. Mitchell, Anisotropy of solar wind turbulence between ion and electron scales. *Phys. Rev. Lett.* **104**, 255002 (2010)
64. C.H.K. Chen, S. Boldyrev, Q. Xia, J.C. Perez, Nature of subproton scale turbulence in the solar wind. *Phys. Rev. Lett.* **110**, 225002 (2013)
65. J.J. Podesta, Evidence of kinetic Alfvén waves in the solar wind at 1 AU. *Solar Phys.* **286**, 529 (2013)

66. L. Matteini, S. Landi, P. Hellinger, F. Pantellini, M. Maksimovic, M. Velli, B.E. Goldstein, E. Marsch, Evolution of the solar wind proton temperature anisotropy from 0.3 to 2.5 AU. *Geophys. Res. Lett.* **34**, L20105 (2007)
67. P. Hellinger, L. Matteini, E. Marsch, Heating and cooling of protons in the fast solar wind between 0.3 and 1 AU: Helios revisited. *J. Geophys. Res.* **116**, A09105 (2011)
68. B.A. Maruca, J.C. Kasper, S.D. Bale, What are the relative roles of heating and cooling in generating solar wind temperature anisotropies? *Phys. Rev. Lett.* **107**, 201101 (2011)
69. G.G. Howes, S.C. Cowley, W. Dorland, G.W. Hammett, E. Quataert, A.A. Schekochihin, A model of turbulence in magnetized plasmas: Implications for the dissipation range in the solar wind. *J. Geophys. Res.* **113**, A05103 (2008)
70. G.G. Howes, A prescription for the turbulent heating of astrophysical plasmas. *Mon. Not. R. Astron. Soc.* **409**, L104–L108 (2010)
71. M.W. Kunz, I.G. Abel, K.G. Klein, A.A. Schekochihin, Astrophysical gyrokinetics: turbulence in pressure-anisotropic plasmas at ion scales and beyond. *J. Plasma Phys.* **84**, 715840201 (2018)
72. S. Galtier, A. Bhattacharjee, Anisotropic weak whistler wave turbulence in electron magnetohydrodynamics. *Phys. Plasmas* **10**, 3065–3076 (2003)
73. J. Cho, A. Lazarian, The anisotropy of electron magnetohydrodynamic turbulence. *Astrophys. J. Lett.* **615**, L41 (2004)
74. S. Boldyrev, J.C. Perez, Spectrum of kinetic-Alfvén turbulence. *Astrophys. J. Lett.* **758**, L44 (2012)
75. T. Passot, P.L. Sulem, A model for the non-universal power law of the solar wind sub-ion-scale magnetic spectrum. *Astrophys. J. Lett.* **812**, L37 (2015)
76. S.S. Cerri, F. Califano, Reconnection and small-scale fields in 2D-3V hybrid-kinetic driven turbulence simulations. *New J. Phys.* **19**, 025007 (2017)
77. L. Franci, S.S. Cerri, F. Califano, S. Landi, E. Papini., A. Verdini, L. Matteini, F. Jenko, P. Hellinger, Magnetic reconnection as a driver for a sub-ion-scale cascade in plasma turbulence. *Astrophys. J. Lett.* **850**, L16 (2017)
78. N.F. Loureiro, S. Boldyrev, Collisionless reconnection in magnetohydrodynamic and kinetic turbulence. *Astrophys. J. Lett.* **850**, L182 (2017)
79. A. Mallet, A. Schekochihin, B.D.G. Chandran, Disruption of Alfvénic turbulence by magnetic reconnection in a collisionless plasma. *J. Plasma Phys.* **83**, 6 (2017)
80. L. Franci, S. Landi, L. Matteini, A. Verdini, P. Hellinger, Plasma beta dependence of the ion-scale spectral break of solar wind turbulence: high-resolution 2D hybrid simulations. *Astrophys. J.* **833**, 91 (2016)
81. D. Vech, A. Mallet, K.G. Klein, J.C. Kasper, Magnetic reconnection may control the ion-scale spectral break of solar wind turbulence. *Astrophys. J. Lett.* **855**, L27 (2018)
82. P. Hellinger, P. Trávníček, J.C. Kasper, A.J. Lazarus, Solar wind proton temperature anisotropy: linear theory and WIND/SWE observations. *Geophys. Res. Lett.* **33**, L09101 (2006)
83. S.D. Bale, J.C. Kasper, G.G. Howes, E. Quataert, C. Salem, D. Sundkvist, Magnetic fluctuation power near proton temperature anisotropy instability thresholds in the solar wind. *Phys. Rev. Lett.* **103**, 211101 (2009)
84. K.G. Klein, G.G. Howes, Predicted impacts of proton temperature anisotropy on solar wind turbulence. *Phys. Plasmas* **22**, 032903 (2015)
85. M.W. Kunz, A.A. Schekochihin, C.H.K. Chen, I.G. Abel, S.C. Cowley, Inertial-range kinetic turbulence in pressure-anisotropic astrophysical plasmas. *J. Plasma Phys.* **81**, 325810501 (2015)
86. C.H.K. Chen, L. Matteini, A.A. Schekochihin, M.L. Stevens, C.S. Salem, B.A. Maruca, M.W. Kunz, S.D. Bale, Multi-species measurements of the firehose and mirror instability thresholds in the solar wind. *Astrophys. J. Lett.* **825**, L26 (2016)
87. M.W. Kunz, J.M. Stone, E. Quataert, Magnetorotational turbulence and dynamo in a collisionless plasma. *Phys. Rev. Lett.* **117**, 235101 (2016)

88. S.S. Cerri, S. Servidio, F. Califano, Kinetic cascade in solar-wind turbulence: 3D3V hybrid-kinetic simulations with electron inertia. *Astrophys. J. Lett.* **846**, L18 (2017)
89. S.S. Cerri, M.W. Kunz, F. Califano, Dual phase-space cascades in 3d hybrid-Vlasov–Maxwell turbulence. *Astrophys. J. Lett.* **856**, L13 (2018)
90. S.S. Cerri, L. Franci, F. Califano, S. Landi, P. Hellinger, Plasma turbulence at ion scales: a comparison between particle in cell and Eulerian hybrid-kinetic approaches. *J. Plasma Phys.* **83**, 705830202 (2017)
91. D. Grošelj, S.S. Cerri, A. Bañón Navarro, C. Willmott, D. Told, N.F. Loureiro, F. Califano, F. Jenko, Fully kinetic versus reduced-kinetic modeling of collisionless plasma turbulence. *Astrophys. J.* **847**, 28 (2017)
92. S.S. Cerri, D. Grošelj, L. Franci, Kinetic plasma turbulence: recent insights and open questions from 3D3V simulations. *Front. Astron. Space Sci.* (in press)
93. J.A. Krommes, The gyrokinetic description of microturbulence in magnetized plasmas. *Annu. Rev. Fluid Mech.* **44**, 175–201 (2012)
94. E.A. Frieman, L. Chen, Nonlinear gyrokinetic equations for low-frequency electromagnetic waves in general plasma equilibria. *Phys. Fluids* **25**, 502–508 (1982)
95. A.J. Brizard, T.S. Hahm, Foundations of nonlinear gyrokinetic theory. *Rev. Mod. Phys.* **79**, 421–468 (2007)
96. G.G. Howes, W. Dorland, S.C. Cowley, G.W. Hammett, E. Quataert, A.A. Schekochihin, T. Tatsumi, Kinetic simulations of magnetized turbulence in astrophysical plasmas. *Phys. Rev. Lett.* **100**, 065004 (2008)
97. G.G. Howes, J.M. TenBarge, W. Dorland, E. Quataert, A.A. Schekochihin, R. Numata, T. Tatsumi, Gyrokinetic simulations of solar wind turbulence from ion to electron scales. *Phys. Rev. Lett.* **107**, 035004 (2011)
98. D. Told, F. Jenko, J.M. TenBarge, G.G. Howes, G.W. Hammett, Multiscale nature of the dissipation range in gyrokinetic simulations of Alfvénic turbulence. *Phys. Rev. Lett.* **115**, 025003 (2015)
99. A. Bañón Navarro, B. Teaca, D. Told, D. Grošelj, P. Crandall, F. Jenko, Structure of plasma heating in gyrokinetic Alfvénic turbulence. *Phys. Rev. Lett.* **117**, 245101 (2016)
100. F. Valentini, F. Califano, P. Veltri, Two-dimensional kinetic turbulence in the solar wind. *Phys. Rev. Lett.* **104**, 205002 (2010)
101. S. Servidio, F. Valentini, F. Califano, P. Veltri, Local kinetic effects in two-dimensional plasma turbulence. *Phys. Rev. Lett.* **108**, 045001 (2012)
102. S. Servidio, K.T. Osman, P. Valentini, D. Perrone, F. Califano, S. Chapman, W.H. Matthaeus, P. Veltri, Proton kinetic effects in Vlasov and solar wind turbulence. *Astrophys. J. Lett.* **781**, L27 (2014)
103. F. Valentini, S. Servidio, D. Perrone, F. Califano, W.H. Matthaeus, P. Veltri, Hybrid Vlasov–Maxwell simulations of two-dimensional turbulence in plasmas. *Phys. Plasmas* **21**, 082307 (2014)
104. S. Servidio, F. Valentini, D. Perrone, A. Greco, F. Califano, W.H. Matthaeus, P. Veltri, A kinetic model of plasma turbulence. *J. Plasma Phys.* **81**, 325810107 (2015)
105. D. Perrone, T. Passot, D. Laveder, F. Valentini, P.L. Sulem, I. Zouganidis, P. Veltri, S. Servidio, Fluid simulations of plasma turbulence at ion scales: comparison with Vlasov–Maxwell simulations. *Phys. Plasmas* **25**, 052302 (2018)
106. O. Pezzi, Y. Yang, F. Valentini, S. Servidio, A. Chasapis, W.H. Matthaeus, P. Veltri, Energy conversion in turbulent weakly collisional plasmas: Eulerian hybrid Vlasov–Maxwell simulations. *Phys. Plasmas* **26**, 072301 (2019)
107. D. Perrone, F. Valentini, S. Servidio, S. Dalena, P. Veltri, Vlasov simulations of multi-ion plasma turbulence in the solar wind. *Astrophys. J.* **762**, 99 (2013)
108. F. Valentini, D. Perrone, S. Stabile, O. Pezzi, S. Servidio, R. De Marco, F. Marcucci, R. Bruno, B. Lavraud, J. De Keyser, G. Consolini, D. Brienza, L. Sorriso-Valvo, A. Retinò, A. Vaivads, M. Salatti, P. Veltri, Differential kinetic dynamics and heating of ions in the turbulent solar wind. *New J. Phys.* **18**, 125001 (2016)

109. L. Sorriso-Valvo, D. Perrone, O. Pezzi, F. Valentini, S. Servidio, I. Zouganelis, P. Veltri, Local energy transfer rate and kinetic processes: the fate of turbulent energy in two-dimensional hybrid Vlasov–Maxwell numerical simulations. *J. Plasma Phys.* **84**, 725840201 (2018)
110. F. Valentini, C.L. Vásconez, O. Pezzi, S. Servidio, F. Malara, F. Pucci, Transition to kinetic turbulence at proton scales driven by large-amplitude kinetic Alfvén fluctuations. *Astron. Astrophys.* **599**, A8 (2017)
111. S.S. Cerri, F. Califano, F. Jenko, D. Told, F. Rincon, Subproton-scale cascades in solar wind turbulence: driven hybrid-kinetic simulations. *Astrophys. J. Lett.* **822**, L12 (2016)
112. A.A. Schekochihin, S.C. Cowley, W. Dorland, G.W. Hammett, G.G. Howes, G.G. Plunk, E. Quataert, T. Tatsuno, Gyrokinetic turbulence: a nonlinear route to dissipation through phase space. *Plasma Phys. Control. Fusion* **50**, 124024 (2008)
113. T. Tatsuno, W. Dorland, A.A. Schekochihin, G.G. Plunk, M. Barnes, S.C. Cowley, G.G. Howes, Nonlinear phase mixing and phase-space cascade of entropy in gyrokinetic plasma turbulence. *Phys. Rev. Lett.* **103**, 015003 (2009)
114. T. Tatsuno, G.G. Plunk, M. Barnes, W. Dorland, G.G. Howes, R. Numata, Freely decaying turbulence in two-dimensional electrostatic gyrokinetics. *Phys. Plasmas* **19**, 122305 (2012)
115. O. Pezzi, S. Servidio, D. Perrone, F. Valentini, L. Sorriso-Valvo, A. Greco, W.H. Matthaeus, P. Veltri, Velocity-space cascade in magnetized plasmas: numerical simulations. *Phys. Plasmas* **25**, 060704 (2018)
116. Y. Kawazura, M. Barnes, A.A. Schekochihin, Thermal disequilibration of ions and electrons by collisionless plasma turbulence. *Proc. Nat. Acad. Sci.* **116**, 771–776 (2019)
117. S. Servidio, A. Chasapis, W.H. Matthaeus, D. Perrone, F. Valentini, T.N. Parashar, P. Veltri, D. Gershman, C.T. Russell, B. Giles, S.A. Fuselier, T.D. Phan, J. Burch, Magnetospheric multiscale observation of plasma velocity-space cascade: hermite representation and theory. *Phys. Rev. Lett.* **119**, 205101 (2017)



Fully Kinetic Simulations: Semi-Lagrangian Particle-in-Cell Codes

6

Patrick Kilian and Felix Spanier

Abstract

This chapter will introduce the particle-in-cell method that connects Eulerian field quantities on a grid with the Lagrangian description of freely moving particles to simulate the evolution of a collisionless plasma. We motivate this description based on the Klimontovich and Maxwell equations, before explaining the basic building block in a numerical implementation. Advanced methods and their trade-offs are mentioned, and good references that contain explanations of implementation details are referenced. The chapter closes with an (incomplete) list of applications for which particle-in-cell codes have successfully been used.

6.1 Theoretical Background of Particle-in-Cell Simulations

6.1.1 Introduction

It has become a common lore that the universe consists of 99% of plasma. While this statement is overly generalized and somewhat misleading, it has to be noted that interplanetary, interstellar, and intergalactic media are in plasma state and all

Supported by Verein zur Förderung der kinetischen Plasmaphysik.

P. Kilian
Space Science Institute, Boulder, CO, USA
e-mail: pkilian@spacescience.org

F. Spanier (✉)
Zentrum für Astronomie, Institut für Theoretische Astrophysik, Universität Heidelberg,
Heidelberg, Germany
e-mail: felix@fspanier.de

interesting sources in space and astrophysics are also dominated by plasmas. It should be noted however that all these plasmas are to some degree “nonthermal,” i.e., their particle distribution has a sizable fraction of particles beyond the Maxwell distribution.

While plasmas with a nonthermal component could be described using numerical standard techniques for hyperbolic conservation laws, the so-called particle-in-cell (PiC) technique has become the standard technique for many plasma simulations.

PiC is a versatile and extremely easy-to-implement kinetic method, which is the main factor behind its popularity. The relatively low computational cost is the next point making PiC such a common choice. Unfortunately this simplicity leads often to carelessness in implementation and simulation design.

This chapter shall give a brief introduction to the physics of kinetic plasmas, the setup of PiC codes, implementation details, and the design of simulations of various physical effects.

6.1.2 Mathematical Description

The definition of a plasma as a quasi-neutral gas with collective behavior covers a wide range of very different physical scenarios. The number of mathematical and numerical descriptions is accordingly very large. An important parameter that determines plasma behavior is the collisionality: In the limit of very high collision frequencies, the plasma is mostly described as a fluid. The opposite limit of no collisions at all is the kinetic regime. The adequate description for this regime is a single-particle description. The collisionality depends on a number of other physical parameters, in particular density, temperature, composition, and degree of ionization. At lower densities collisions are less common, and long-range electromagnetic forces between charged particles play a (comparatively) larger role. At high temperatures particles have larger relative velocities and spend less time close to each other which reduces the influence of collisions. This can be seen from the v^{-3} dependence of the Coulomb cross section. A low degree of ionization will make collisions with the abundant neutrals common and important where a highly ionized plasma with the same density of charged particles but without the neutrals may very well be effectively collisionless. There is a (possibly exceedingly tiny) fraction of charged particles in any gas of finite temperature. As the temperature increases to the point where the average energy per particle crosses the ionization threshold, charged particles become much more abundant. In most cases that are studied with particle-in-cell simulations, the degree of ionization (either due to high temperatures or other ionizing mechanisms) is so large that neutrals will be completely ignored. But in principle the fraction of ionized particles is going to vary between zero and unity, without attaining either limit or without a first-order phase transition. As such the often employed “plasma as the fourth state of matter” is technically not correct. But in the following, we will not concern ourselves with such disputes of theoretical purity and will instead consider a plasma with abundant charged particles, sufficient temperature, and negligible collisions, similar to the

choice in Chap. 4. If collisions (that are in some sense weak) are of interest, they can be accommodated by, e.g., the algorithms described by Takizuka and Abe [1] or Miller and Combi [2].

A kinetic plasma could be described by taking all plasma particles and their pairwise forces into account (this would equal an N-body code). This approach is however neither feasible nor sensible: The number of particles N is extremely large, and the runtime is $\mathcal{O}(N^2)$ at worst or $\mathcal{O}(N \log N)$ at best.

The particle-in-cell (PiC) approach avoids the use of particle-particle forces and their inherent computational need. Instead the interaction of particles with the electromagnetic field generated by the moving charges is considered. Additionally, particles are not tracked individually, but bunches of particles are combined into a single computational macroparticle. The description of particle bunches generating fields and interacting with them is somewhat naïve. From a mathematical viewpoint, the PiC algorithm uses phase-space sampling to solve the Vlasov equation (or the Maxwell-Vlasov system).

The Vlasov equation can be derived easily from a set of particles through the Klimontovich equation. Starting with the microscopic distribution of a species α in the six-dimensional phase space

$$N_\alpha(\mathbf{x}, \mathbf{v}, t) = \sum_{k=1}^{N_0} \delta(\mathbf{x} - \mathbf{x}_k(t)) \delta(\mathbf{v} - \mathbf{v}_k(t)) \quad (6.1)$$

Here each particle k of the N_0 particles of species α has discrete positions $\mathbf{x}_k(t)$ and velocities $\mathbf{v}_k(t)$ in phase space. The particles with charge q_α and m_α are affected by electromagnetic fields, and their motion is governed (in the non-relativistic limit) by the set of differential equations (in CGS)

$$\frac{d\mathbf{x}_k(t)}{dt} = \mathbf{v}_k(t), \quad \frac{d\mathbf{v}_k(t)}{dt} = \frac{q_\alpha}{m_\alpha} (\mathbf{E}^m(\mathbf{x}_k(t), t) + \mathbf{v}_k(t)/c \times \mathbf{B}^m(\mathbf{x}_k(t), t)) \quad (6.2)$$

The Klimontovich equation can then be derived by taking the time derivative of Eq. (6.1) and using Eq. (6.2).

$$\frac{\partial N_\alpha(\mathbf{x}, \mathbf{v}, t)}{\partial t} = \frac{\partial}{\partial t} \sum_{k=1}^{N_0} \delta(\mathbf{x} - \mathbf{x}_k(t)) \delta(\mathbf{v} - \mathbf{v}_k(t)) \quad (6.3)$$

$$\begin{aligned} &= -\mathbf{v} \cdot \frac{\partial N_\alpha(\mathbf{x}, \mathbf{v}, t)}{\partial \mathbf{x}} \\ &\quad - \frac{q_\alpha}{m_\alpha} (\mathbf{E}^m(\mathbf{x}(t), t) + \mathbf{v}(t)/c \times \mathbf{B}^m(\mathbf{x}(t), t)) \cdot \frac{\partial N_\alpha(\mathbf{x}, \mathbf{v}, t)}{\partial \mathbf{v}} \end{aligned} \quad (6.4)$$

This can be rewritten as

$$\frac{\partial N_\alpha}{\partial t} + \mathbf{v} \cdot \frac{\partial N_\alpha}{\partial \mathbf{x}} + \frac{q_\alpha}{m_\alpha} (\mathbf{E}^m + \mathbf{v}/c \times \mathbf{B}^m) \cdot \frac{\partial N_\alpha}{\partial \mathbf{v}} = 0 \quad (6.5)$$

This is the Klimontovich equation that describes the microphysical distribution function. In conjunction with the Maxwell equations, this equation describes a collisionless plasma completely. Given that N_0 is large, the Klimontovich equation is unusable for any practical purpose.

A better description is achieved by going to the ensemble average, which includes all physically identical realizations of the plasma. The phase space density of the ensemble average is given by

$$f_\alpha(\mathbf{x}, \mathbf{v}, t) = \langle N_\alpha(\mathbf{x}, \mathbf{v}, t) \rangle \quad (6.6)$$

The microscopic quantities of N_α , \mathbf{B}^m , and \mathbf{E}^m can be written as a sum of ensemble average and perturbation:

$$N_\alpha(\mathbf{x}, \mathbf{v}, t) = f_\alpha(\mathbf{x}, \mathbf{v}, t) + \delta N_\alpha(\mathbf{x}, \mathbf{v}, t) \quad (6.7)$$

$$\mathbf{E}^m(\mathbf{x}, \mathbf{v}, t) = \mathbf{E}(\mathbf{x}, \mathbf{v}, t) + \delta \mathbf{E}^m(\mathbf{x}, \mathbf{v}, t) \quad (6.8)$$

$$\mathbf{B}^m(\mathbf{x}, \mathbf{v}, t) = \mathbf{B}(\mathbf{x}, \mathbf{v}, t) + \delta \mathbf{B}^m(\mathbf{x}, \mathbf{v}, t) \quad (6.9)$$

We can substitute the definitions in Eqs. (6.7)–(6.9) into Eq. (6.5) and take the ensemble average. Making use of the linearity of the averaging to split the equation into terms that can be averaged individually and applying definition (6.6), we obtain

$$\begin{aligned} & \frac{\partial f_\alpha(\mathbf{x}, \mathbf{v}, t)}{\partial t} + \mathbf{v} \cdot \frac{\partial f_\alpha(\mathbf{x}, \mathbf{v}, t)}{\partial \mathbf{x}} \\ & + \frac{q_\alpha}{m_\alpha} (\mathbf{E}(\mathbf{x}(t), t) + \mathbf{v}(t)/c \times \mathbf{B}(\mathbf{x}(t), t)) \cdot \frac{\partial f_\alpha(\mathbf{x}, \mathbf{v}, t)}{\partial \mathbf{v}} \\ & = -\frac{q_\alpha}{m_\alpha} \left\langle (\delta \mathbf{E}^m(\mathbf{x}(t), t) + \mathbf{v}(t)/c \times \delta \mathbf{B}^m(\mathbf{x}(t), t)) \cdot \frac{\partial \delta N_\alpha(\mathbf{x}, \mathbf{v}, t)}{\partial t} \right\rangle \end{aligned} \quad (6.10)$$

The term on the RHS is the ensemble average of a second-order product of perturbations in different realizations of a plasma state. This quantity is negligibly small. The LHS of the equation then yields the Vlasov equation

$$\frac{\partial f_\alpha}{\partial t} + \mathbf{v} \cdot \frac{\partial f_\alpha}{\partial \mathbf{x}} + \frac{q_\alpha}{m_\alpha} (\mathbf{E} + \mathbf{v}/c \times \mathbf{B}) \cdot \frac{\partial f_\alpha}{\partial \mathbf{v}} = 0 \quad (6.11)$$

From the viewpoint of physics, the Vlasov equation is the representation of the time evolution of the phase space distribution ensemble average (in combination with Maxwell's equations). Mathematically the Vlasov equation is a hyperbolic

conservation equation (which corresponds to the fact that the phase space volume is conserved).

Formally a large set of methods exist to solve a set of hyperbolic conservation laws (even for the six-dimensional phase space), and a number of solutions for lower-dimensional systems has been implemented.

Direct numerical simulation is severely limited: In the first place, it requires a solution on a six-dimensional grid, and on the other hand, the requirement to have constantly positive phase space density leads to strong diffusion. These two problems together cause a massive computational overhead that only in very special cases justifies the use of Vlasov methods. One such case is when low numerical noise is needed.

Rather than solving the Vlasov equation for the full distribution function, samples of the phase space can be used. As Liouville's theorem states, the volume of these phase space samples is constant. The motion of the phase space volumes can be traced as single particles. However the phase space samples do not represent individual physical particles, but rather a bunch of w_p physical particles with (nearly) identical position and velocity that stay together during time evolution. This number w_p is also known as the statistical weight of the phase space sample or the macrofactor.

The distinction made here is between the Eulerian and Lagrangian pictures: The Vlasov equation as hyperbolic conservation law can be treated from the viewpoint of the change of the distribution function at fixed points (i.e., on a fixed grid of the phase space) or by following the motion of single phase space points. The latter is the so-called Lagrangian picture, while the former is the Eulerian picture. This distinction is not limited to the Vlasov equation or the phase space. It is a common concept used also in hydrodynamics, where fluid elements or the fluid velocities on a grid could be traced.

There are still different possibilities for the implementation of a Lagrangian picture: As discussed above there are PiC and N-body codes. The N-body method is a particle-particle method, where the particles are treated in a Lagrangian way and the inter-particle forces are calculated directly. The PiC method on the other hand is a particle-mesh (PM) method, where particles are Lagrangian, but their inter-particle forces are calculated via a grid, which stores fields.

Following the phase space samples, using the PiC algorithm is obviously a solution to the Vlasov equation. As a numerical solution for phase space samples, it is obviously not an exact solution of the Vlasov equation. But besides unavoidable problems of numerical methods (e.g., limited accuracy), the PiC method has two distinct problems: noise from phase space sampling and discretization errors from the limited resolution of the grid.

The mathematical viewpoint is clear: PiC is a method, where the phase space distribution is sampled and forces are calculated using a particle-to-mesh method. This in turn yields solutions, which are solutions to the Vlasov equation, i.e., collisionless plasmas. There are no further restrictions which kind of plasma may be described using the PiC method, but for practical purposes, only small-scale simulations are possible.

6.2 Numerical Implementation

Particle-in-cell codes are semi-Lagrangian in that they follow a Lagrangian description of phase space samples represented as macroparticles coupled to electromagnetic fields described in an Eulerian way on a grid. Connecting the two parts correctly is crucial in the numerical implementation, but let us start with the relatively simple step of evolving the electric and magnetic fields on the grid for one time step. The simulation grid can resolve one, two, or three spatial dimensions but typically vector quantities such as field vectors or particle velocities retain all three components.

6.2.1 Field Solvers

This subsection discusses how electromagnetic fields are calculated from source terms (moments of the distribution function of all charged particle species). The sources are assumed to be known on a uniform grid of rectangular cells. How the source terms are calculated is discussed in Sect. 6.2.4. Non-rectangular grids are discussed in the review [3].

Particle-in-cell codes commonly use a finite-difference formulation based on the electromagnetic fields. Codes that use the scalar electrostatic potential and magnetic vector potential also exist [4, 5] but are less common. A finite element view of the fields was considered in [6] but leads to a non-local matrix equation for the fields. Lumping spatial and temporal averaging operators reduces the scheme to the same expressions that are obtained from the finite-difference formulation. The typical arrangement of field samples in this finite-difference formulation is given by the Yee lattice described in [7]. Components of the electric field E_i are located at the midpoints of edges pointing in coordinate direction x_i , whereas magnetic flux components B_i are located at the center of cell faces which have normal vectors pointing in direction x_i , as shown in Fig. 6.1a. In principle the two dual subgrids could be interchanged, but this arrangement neatly follows the view of electromagnetism as a discrete exterior calculus, where the electric field is a one-form that naturally lives on cell edges. The magnetic flux is represented by two-forms assigned to cell faces. This view of computational electrodynamics is rather recent but has proven useful when designing field solvers on unstructured grids [8, 9].

The staggered arrangement of components is non-trivial, and it looks as if interpolation is needed to calculate derivatives at the correct positions. But the contrary is true as we will see in a moment. We start our consideration from Maxwell's equation in CGS units. These are still widely used in particle-in-cell codes as they make the magnitude of electric and magnetic components similar. (Of course codes in SI units such as EPOCH [10] or codes that treat the problem completely in dimensionless units such as VPIC [11, 12] also exist.) Since we treat all material effects due to the plasma explicitly by having particles react to the fields

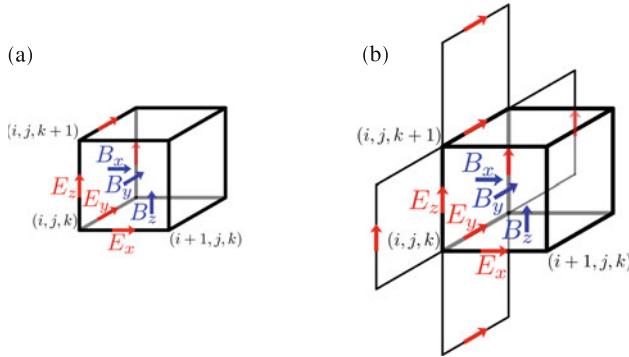


Fig. 6.1 Second- and fourth-order finite-difference time-domain schemes. **(a)** Finite difference approximation in Yee scheme. **(b)** Finite difference approximation in simplest fourth-order scheme

E and **H**, we can use them interchangeably with the respective flux densities **D** and **B**:

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} \quad (6.12)$$

$$\nabla \times \mathbf{B} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \mathbf{J} \quad (6.13)$$

$$\nabla \cdot \mathbf{E} = 4\pi \rho \quad (6.14)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (6.15)$$

Gauss's law in Eq. (6.14) is the corner stone for electrostatic particle-in-cell codes, but we will ignore it for now. Its equivalent for the magnetic field, Eq. (6.15), that describes the absence of monopoles is mostly relevant as an initial condition. Faraday's law and Ampere's law can be rewritten as evolution equations:

$$\frac{\partial \mathbf{B}}{\partial t} = -c \nabla \times \mathbf{E} \quad (6.16)$$

$$\frac{\partial \mathbf{E}}{\partial t} = c \nabla \times \mathbf{B} - 4\pi \mathbf{J} \quad (6.17)$$

The time evolution given by these equations will not violate Eqs. (6.14) and (6.15) if they were satisfied initially. Equations (6.16) and (6.17) imply that the curl of the electric field does not have to be known at the location of the electric field, but rather at the location of the dual, magnetic field.

The classical solution to store electric and magnetic fields on the grids that yields the curl of the fields at the required position is the Yee grid. Here, e.g., B_y is stored at cell face $(i + 1/2, j, k + 1/2)$, while E_y is stored at $(i, j + 1/2, k)$.¹ Not only does this process result in curls calculated at the correct position without interpolation, but the curl is calculated in the second order immediately.

The temporal derivative in Eqs. (6.16) and (6.17) can be solved using the second-order spatial derivatives. The customary solution is to leap-frog the fields, i.e., use the curl of the electric field at time $t \Delta t$ to update the magnetic field from time $(t - 1/2)\Delta t$ to $(t + 1/2)\Delta t$, and then use the curl of the magnetic field at that time to update the electric field to $(t + 1)\Delta t$. This leads to a scheme that is also second order accurate in time. The time step Δt cannot be larger than the CFL limit $\Delta t < \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}/c$ to avoid instability (see more below).

To update the particle position and location, it is useful to have electric and magnetic fields at the position of the particle at the same time. The solution to this is to split the update of the magnetic field and use

$$\mathbf{B}^t = \mathbf{B}^{t-1/2} - \frac{1}{2}c \cdot \Delta t \cdot \nabla \times \mathbf{E}^t \quad (6.18)$$

$$\mathbf{B}^{t+1/2} = \mathbf{B}^t - \frac{1}{2}c \cdot \Delta t \cdot \nabla \times \mathbf{E}^t \quad (6.19)$$

$$\mathbf{E}^{t+1} = \mathbf{E}^t + c \cdot \Delta t \cdot \nabla \times \mathbf{B}^{t+1/2} - 4\pi \Delta t \cdot \mathbf{j}^{t+1/2} \quad (6.20)$$

This solver is the most common method and is accurate to second order with respect to a given current.

Since small numbers of macroparticles lead to an increased noise, solutions are sought after to remove the noise. Increasing particle numbers is typically not an option as the noise level decreases as $1/\sqrt{N}$. For the cause of the noise and properties thereof, see, for example, [13] and [14, Appendix D]. Filtering can be applied to either the source term $\mathbf{j}^{t+1/2}$ to avoid the production of noise or the fields themselves to dampen existing noise. The latter is possible since noise is concentrated at high k values and numerical errors (e.g., dispersion errors) are largest there. This low-pass filtering is described in [15] and changes Eqs. (6.18) and (6.19).

The stability condition CFL, named after Courant, Friedrichs and Levy, is mentioned many times in this book. For its better understanding let's take a quick look at this important condition, common for explicit schemes. For how to avoid it in implicit schemes, see Chap. 8. For simplicity let's analyze the case where there is no current \mathbf{j} and where fields can vary only in x and t . In that case Eqs. (6.18)–(6.20)

¹ The grid cells have edge length Δx , Δy , and Δz , and any position can be given as $\mathbf{x} = (i \cdot \Delta x, j \cdot \Delta y, k \cdot \Delta z)$.

for E_y and B_z reduce to

$$B_z^{t+1/2} \left(i + \frac{1}{2} \right) = B_z^{t-1/2} \left(i + \frac{1}{2} \right) - c \cdot \Delta t \cdot \frac{(E_y^t(i+1) - E_y^t(i))}{\Delta x}, \quad (6.21)$$

$$E_y^{t+1}(i) = E_y^t(i) + c \cdot \Delta t \cdot \frac{\left(B_z^{t+1/2} \left(i + \frac{1}{2} \right) - B_z^{t+1/2} \left(i - \frac{1}{2} \right) \right)}{\Delta x} \quad (6.22)$$

Assuming that there is no initial $B_z^{t-1/2}$ and that E_y^t has the form $A \sin(kx)$, we find

$$E_y^{t+1} = \frac{c^2 \Delta t^2}{\Delta x^2} \cdot \frac{1 - \cos(k \Delta x)}{2} \cdot A \sin(kx) \quad (6.23)$$

So for the most unstable $k = \pi/\Delta x$ —which is also the highest k that can be supported by the grid—the amplitude has grown after a single time step by a factor $c^2 \Delta t^2 / \Delta x^2$. If this factor is larger than unity, the time advance will amplify waves exponentially with time. To avoid that we require $\Delta t \leq \Delta x/c$. Intuitively this requirement say “the time step should be short enough that waves don’t travel further than one grid cell per step,” which makes sense since our numerical derivatives only consider neighboring cells. The simplified analysis shown here can be generalized to allow propagation also in y and z in which case we will find the condition given above. In other plasma models, other waves might be the most unstable modes, e.g., Whistler waves in many hybrid-kinetic codes or fast-magneto-sonic waves in MHD, typically because the light wave considered here is not present in those models. Details of the discretization of the spatial derivatives will introduce a numerical constant not too different from unity. Additional details as well as stability limits that are introduced by the presence of the particles can be found in standard textbooks such as [16, 17].

The one instability that plagues electromagnetic particle-in-cell codes and is the main driver to go beyond the simple and robust field solver described above is the numerical Cherenkov instability (NCI). This long known instability (see [18]) occurs when particles move close to the speed of light and couple to light waves. In reality this only happens when the influence of a medium reduces the speed of light below the vacuum value and particle can reach the local phase velocity. Numerically this can happen when the numerical scheme introduces a dispersion error which reduces the speed of light. The Maxwell solver presented above does that, especially for short wavelength comparable to the length of the cell edges.

In other applications of finite-difference time-domain (FDTD) schemes, the solution would typically be to use a higher-order approximation of the discrete curl operator. Figure 6.1b shows how the straightforward extension to the fourth order in

space would look like. The right-hand side of Eq. (6.16) now reads

$$\begin{aligned}
 (\nabla \times \mathbf{E})_x(i, j, k) \approx & \frac{27}{24} \frac{E_y(i, j, k) - E_y(i, j, k+1)}{\Delta z} \\
 & - \frac{E_y(i, j, k-1) - E_y(i, j, k+2)}{\Delta z} \\
 & + \frac{27}{24} \frac{E_z(i, j+1, k) - E_z(i, j, k)}{\Delta y} \\
 & - \frac{E_z(i, j+2, k) - E_z(i, j-1, k)}{\Delta y}
 \end{aligned} \tag{6.24}$$

in this case. Also the temporal derivatives can be approximated in the fourth order [19]. For typical PiC simulations with marginally stable time steps, the gain in accuracy is negligible compared to the computational cost.

The way forward has therefore been to use schemes that are still “only” second order but with a smaller constant in front of the error term. Many such non-standard finite-difference (NSFD) schemes have been investigated over the years.

One such scheme, proposed by Hadi and Piket-May [20] under the name “M24” makes use of the grid locations used by the standard Yee scheme, the locations used by the fourth-order scheme, and additional locations that close a larger loop around the midpoint. The relative weights of the three loops are in principle arbitrary, subject to the constraint that they have to sum to one. The remaining two free parameters can be tuned to optimize the dispersion properties of the scheme. The original paper suggests some values, but more interesting is [21] that shows how these parameters can be chosen to reduce NCI.

So far all schemes only used grid locations at $i\Delta x$, but [22] shows that using $(i\pm 1)\Delta x$ can be useful, at least for a narrow range of frequencies propagating on the grid. Subsequently [23] extended that to broad bandwidth. Figure 6.2b illustrates all the grid locations that are used in this family of schemes. Choosing different relative weights between the different grid locations allows to tune for different optimization goals. An overview is given in [24].

A different approach is the pseudo-spectral solution of the field equations. As known from, e.g., hydrodynamics, the partial differential equations in real space are converted to ordinary equations in Fourier space. This eliminates finite stencils but requires a global FFT [25, 26]. For constant current density, Maxwell’s equations can be integrated analytically in time leading to pseudo-spectral analytic time-domain (PSATD) schemes that are completely free of numerical dispersion. Since global parallel FFTs require a lot of communication and can limit scalability, there have been some attempts to avoid them and instead use local FFTs to build field solvers [27].

For the limit of very small particle velocities (as compared to the speed of light) it might be possible to remove light waves completely from the plasma model and to reduce computational expense. The straightforward way to do this is the

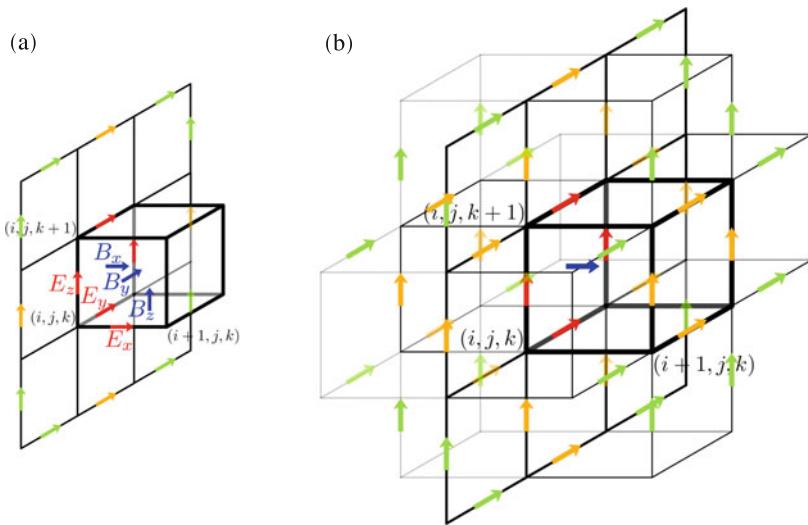


Fig. 6.2 Non-standard finite-difference time-domain schemes. **(a)** Finite difference approximation in M24 scheme. **(b)** Finite difference approximation in Cole-Karkkainen schemes

electrostatic plasma model, where the electric field is not updated using an evolution equation based on Ampere's law Eq. (6.17), but is directly calculated from Gauss's law Eq. (6.14). This of course limits the electric field to its longitudinal, curl-free component, and magnetic fields are no longer time dependent. Intuitively that makes sense as very slow particle motion implies very small currents. Historically some of the first particle-in-cell codes were electrostatic (see, e.g., [28, 29]). Gauss's law can be solved in higher dimensions using either spectral solvers [17] or multigrid solvers.

Besides the domain of electrostatic and electromagnetic models, there is another intermediate regime called Darwin approximation [30] which includes magnetodynamic effects up to order $O((v/c)^2)$ but which does not consider light waves. This is equivalent to removing the divergence-free part of the displacement current. While time resolution constraints are relaxed since the speed of light is no longer present, the differential equations representing the electromagnetic field are now elliptic requiring a global field solver.

6.2.2 Interpolation

The previous section only discussed Eulerian quantities on the regular grid. But the electromagnetic fields act on the Lagrangian particles, which requires knowledge of the field strength at the location of the particle, not just at the regular grid locations. This crucial switch in view point is done by means of an interpolation that we will discuss next.

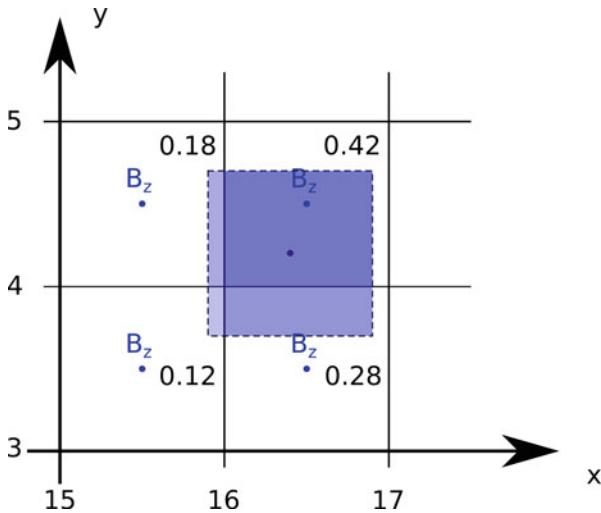


Fig. 6.3 Interpolation using the overlap between the cell and the particle, weighted by the shape function

Figure 6.3 illustrates how the interpolation from grid quantities (magnetic field B_z in this case) to the particle location is done. As discussed in the last section, the numerical macroparticle represents a sample of the phase space with finite extent and a macrofactor or (statistical) weight w_p given by the number of physical particles it represents. The fraction of the phase space volume that is located in the cell is used as an interpolation weight for the field strength contribution from that cell. Contributions from different cells are then added up to compute the mean magnetic field strength at the particle location. In a single dimension, this calculation can be expressed by

$$\bar{B}_z(x_p) = \sum_i B_z((i + 1/2)\Delta x) \cdot \int_{i\Delta x}^{(i+1)\Delta x} S(x - x_p) dx \quad (6.25)$$

Since this calculation is required for all particles and all field components in each time step, the integral over the shape function S is not performed numerically, but rather done analytically ahead of time to obtain the weight function W :

$$W(x) = \int_{-\Delta x/2}^{\Delta x/2} S(x - x') dx' \quad (6.26)$$

Using this and suppressing the explicit location of the gridded quantity in the cell, we can write the two-dimensional interpolation as

$$\bar{B}_z(\mathbf{x}_p) = \sum_{i,j} B_z(i, j) \cdot W(x_p - (i + 1/2)\Delta x) \cdot W(y_p - (j + 1/2)\Delta y) \quad (6.27)$$

Many different choices for the shape function S are possible, and different codes make different design choices here. B-Splines are commonly chosen, but is not the only family of function that satisfies the following requirements: The shape function should not include the statistical weight of the macroparticle, but rather should be normalized, i.e., $\int_{-\infty}^{\infty} S(x) dx = 1$, to ensure that the plasma density is correctly represented. The shape function should be symmetric, i.e., $S(-x) = S(x)$, to avoid unphysical self-forces on the particles. To allow efficient evaluation of Eq. (6.27), the shape function should have compact support, i.e., be non-zero only up to some small x_{max} to limit the sum to only go over a few grid cells.

The *nearest grid point* or NGP scheme is the minimal example satisfying all requirements: The shape function is the δ -function $S(x) = \delta(x)$. This choice is simple to implement and computationally cheap, but the field that particles see is non-continuous and the current induced by particles using this shape function switches from one cell to another when the interface between those cells is crossed.

The *cloud-in-cell* or CIC algorithm suggested by Birdsall and Fuss [31] was a major step forward in the performance and usefulness of particle-in-cell codes. CIC views each particle as a cloud of constant density and with the same size as a grid cell but freely movable relative to the grid. The shape and weight function that generate that view are given by

$$S_{\text{CIC}}(x) = \begin{cases} 1 & |x| \leq 1/2 \\ 0 & \text{else} \end{cases}, \quad W_{\text{CIC}}(x) = \begin{cases} 1 - |x| & |x| \leq 1 \\ 0 & \text{else} \end{cases} \quad (6.28)$$

and illustrated in Fig. 6.4 (top left).

This simple shape function is not only a major step forward, but has the added benefit of generating an entire family of shape function of increasing order, by using the weight function W for the previous order as the shape function S of the new, higher order. The next step after CIC in this chain is called *triangular-shaped cloud* or TSC. This shape function has the advantage of being continuous. This implies that the Fourier transform of it falls faster toward higher k . In [32] it has been shown that this decreases aliasing of high k . This in turn decreases the finite grid instability—a numerical artifact that occurs, when the numerical grid does not resolve the Debye length and that artificially heats the plasma. The use of higher-order shape functions improves the conservation of thermal energy[33].

Higher-order shape function from the same family such as *piecewise quadratic shape* (PQS) illustrated in Fig. 6.4 (top right) has been used. Higher orders have an increasing number of continuous derivatives and therefore quicker drop-off in Fourier space but at the cost of each particle interacting with an increasing number of cells. The smoother fields generated by higher-order shape functions reduce unphysical binary forces between nearby particles and therefore numerical collisionality [34, 35]. Otherwise orders beyond TSC are not used widely enough that a consensus as to their names has emerged.

So far only dimension-by-dimension interpolation has been discussed. This construction leads to codes that conserve total momentum, if the shape functions

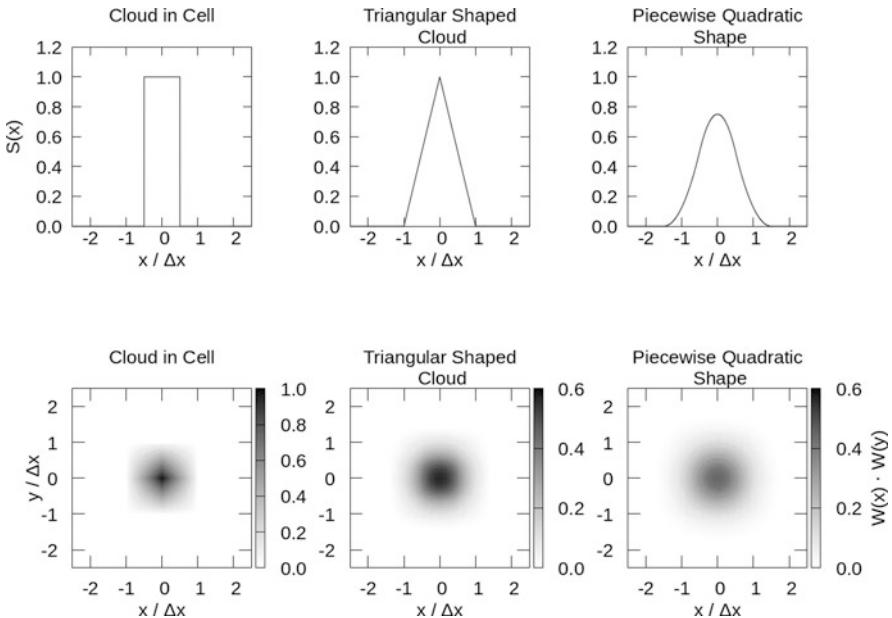


Fig. 6.4 Different shape functions in 1D and 2D

used during interpolation and deposition (see below in Sect. 6.2.4) are identical. Alternatively it is possible to combine shape function of different orders to obtain a code that conserves total energy [36]. A good example of this is the VPIC code (description in [11]) that mixes NGP and CIC shape functions. Magnetic field components B_i are interpolated using linear interpolation in the x_i coordinate and NGP in the other two coordinates. Electric field components E_i use NGP in the x_i coordinate and CIC for the other two directions. The low-level interpolation unfortunately requires very high particle numbers to reach acceptable noise levels.

Since many different choices for shape functions are possible, many comparisons have been made [37–39]. No universal winner has emerged, and it is advisable to test different shape functions to make sure that physical results are insensitive to the choices made. A recent idea by Wu and Qin [40] is to use adaptive width per particle and update them during the course of the simulation to reduce overall noise.

6.2.3 Particle Motion

Once the electromagnetic fields have been interpolated to the location of the particle, it is necessary to update the position and velocity of the particle accordingly. Since the first publication of the particle push by Boris [41], basically all codes use this method, possibly with minor adjustments.

Similar to the leap-frog procedure for the electromagnetic fields, the particle's position and velocity are updated according to

$$\mathbf{v}(t + \Delta t/2) = \mathbf{v}(t - \Delta t/2) + \mathbf{F}(x(t), v(t)) \cdot \Delta t / m \quad (6.29)$$

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t + \Delta t/2) \cdot \Delta t \quad (6.30)$$

When the force depends only on position, this can be solved straightforwardly. For velocity-dependent forces—magnetic fields are the most important case—there is a dependence on $\mathbf{v}(t)$, where the velocity is not known at time t . The interpolation $\mathbf{v}(t) \approx 1/2 (\mathbf{v}(t - \Delta t/2) + \mathbf{v}(t + \Delta t/2))$ solves this problem to the second order, but this leads to an implicit dependence of Eq. (6.29) on $\mathbf{v}(t + \Delta t/2)$ on both sides.

The Boris push allows us to solve this directly and efficiently. The computational efficiency is important since typical simulations contain many more particles than grid cells.

In order to account for relativistic versions, the formulation shall be done using relativistic momentum per rest mass $\mathbf{u} = \mathbf{p}/m = \gamma \mathbf{v}$. Here γ is the Lorentz factor of the macroparticle. This quantity does not suffer from the problem that it has a fixed upper value, and thus we avoid the loss of precision due to the finite resolution of floating point numbers close to the speed of light. Equations (6.29) and (6.30) then read

$$\mathbf{u}(t + \Delta t/2) = \mathbf{u}(t - \Delta t/2) + \mathbf{F}(x(t), v(t)) \cdot \Delta t / m \quad (6.31)$$

$$\gamma(t + \Delta t/2) = \sqrt{1 + \frac{\mathbf{u}(t + \Delta t/2) \cdot \mathbf{u}(t + \Delta t/2)}{c^2}} \quad (6.32)$$

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{u}(t + \Delta t/2) \cdot \Delta t / \gamma(t + \Delta t/2) \quad (6.33)$$

$$\mathbf{F}(x(t), v(t)) = q (\mathbf{E}(x(t)) + \mathbf{v}(t)/c \times \mathbf{B}(x(t))) \quad (6.34)$$

One important thing to note is that the macrofactor w_p cancels between Eqs. (6.31) and (6.34), since both charge q and mass m of a macroparticle are increased compared to q_α and m_α of a single physical particle. Positions \mathbf{x} and velocities \mathbf{v} are both unscaled. This means that Eqs. (6.31)–(6.34) apply to physical particles and phase space samples in the same way and the dynamics of macroparticles in the code can be intuitively interpreted as particle trajectories.

The major insight that allows to solve the implicit equations of motion is that the magnetic field can only change the direction of the velocity vector. It is therefore possible to construct a suitable rotation using geometric considerations (Fig. 6.5). The effect of the electric field is purely an acceleration in the direction of the electric field. To maintain reversibility in time and second-order accuracy, it is split into two equal parts performed before and after handling the magnetic field. Using this insight we can perform the following steps:

$$\mathbf{u}_1 = \mathbf{u}(t - \Delta t/2) + \frac{q}{m} \mathbf{E}(t) \frac{\Delta t}{2} \quad (6.35)$$

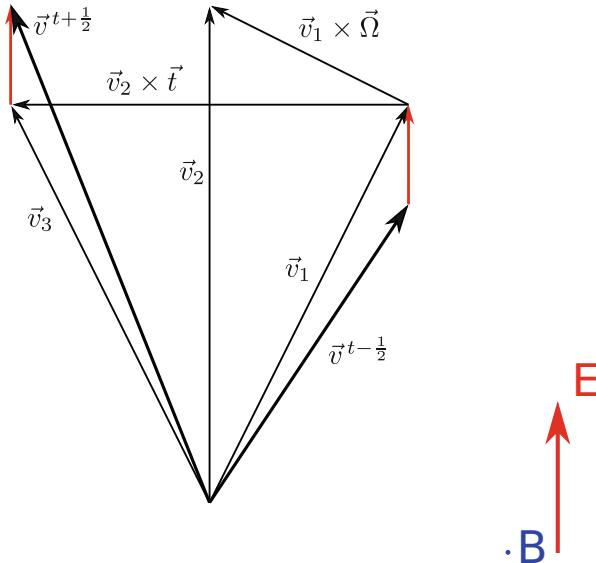


Fig. 6.5 Velocity diagram of the Boris push: The red arrows indicate the change in velocity $q/m\mathbf{E}(t)\Delta t/2$ from the acceleration by the electric field $E(t)$ at the beginning and end of the Boris push. The vector Ω is a normalized magnetic field \mathbf{B} such that a cross product with it produces half the final rotation. This allows the construction of \mathbf{v}_2 and the final accurate solution

$$\Omega = \tan \left(\frac{q}{\gamma mc} |\mathbf{B}(t)| \frac{\Delta t}{2} \right) \frac{\mathbf{B}(t)}{|\mathbf{B}(t)|} \quad (6.36)$$

$$\mathbf{t} = \frac{2}{1 + \Omega \cdot \Omega} \Omega \quad (6.37)$$

$$\mathbf{u}_2 = \mathbf{u}_1 + \mathbf{u}_1 \times \Omega \quad (6.38)$$

$$\mathbf{u}_3 = \mathbf{u}_1 + \mathbf{u}_2 \times \mathbf{t} \quad (6.39)$$

$$\mathbf{u}(t + \Delta t/2) = \mathbf{u}_3 + \frac{q}{m} \mathbf{E}(t) \frac{\Delta t}{2} \quad (6.40)$$

It is worth noting that the original algorithm by Boris contains the *tan* in Eq. (6.36) and the suggestion to expand it at least up to fourth order if it is too expensive to compute *tan* to machine precision. However many codes since then have made the approximation $\Omega \approx q/(\gamma mc) \mathbf{B}(t) \Delta t/2$. This produces only the correct gyroradius, not the correct gyroangle.

The Lorentz factor γ in Eq. (6.36) is approximated by $\gamma = \sqrt{1 + \mathbf{u}_1 \cdot \mathbf{u}_1/c^2}$, which according to [42] can lead to spurious forces. Other modifications of the Boris push, such as [43] and [44], exist as well. Some comparison between different variants can be found in [45] or [46].

6.2.4 Deposition and Filtering

Now that the electromagnetic field has acted on the particle, we need to calculate—in the case of an electrostatic code—the charge density distribution ρ resulting from the new locations of the particles or—in the case of an electromagnetic code—the current density distribution \mathbf{j} . This requires interpolation from the particle locations back to the grid and is typically the most expensive operation in the simulation code. As [17] shows it is desirable to use the same interpolation method for both the interpolation from the grid to the particle location and the deposition of the particle quantities ρ and \mathbf{j} onto the grid to avoid spurious self-forces.

The deposition of $\rho(t + \Delta t)$ based on the newly calculated $\mathbf{x}(t + \Delta t)$ is straightforward and simply given by

$$\rho(i\Delta x, j\Delta y) = \sum_p q_p W(x_p - i\Delta x) \cdot W(y_p - j\Delta y) \quad (6.41)$$

The sum is taken over all particles, multiplying their charge q_p (which is given by $w_p \cdot q_\alpha$ for a macroparticle with macrofactor w_p , representing w_p physical particles of species α with charge q_α each) with the weight function based on the distance between the particle position $\mathbf{x}_p(t + \Delta t)$ and the origin of the grid cell, where the charge density is located. One might be tempted to locate the charge density at the center of the cell, to allow for the nice interpretation as the charge contained in the cell, but typically the field solver calculates the electrostatic potential at the location where the charge density is located. Picking the corners of the cell allows to use a straightforward finite difference approximation to calculate the electric field at the midpoint of the edges.

Calculating the current $\mathbf{j}(t + \Delta t/2)$ based on the old location $\mathbf{x}(t)$, the new location $\mathbf{x}(t + \Delta t)$ and the velocity $\mathbf{v}(t + \Delta t/2)$ is slightly more complicated.

The current carried by a particle with charge q_p moving at velocity $\mathbf{v}(t + \Delta t/2)$ is given by $\mathbf{j} = q_p \mathbf{v}(t + \Delta t/2)$, and it is located around the position at this time, $\mathbf{x}(t + \Delta t/2)$, that can be approximated by $(\mathbf{x}(t) + \mathbf{x}(t + \Delta t))/2$. Figure 6.6 illustrates that this current can simply be interpolated onto the grid using the weight functions, just as for the charge density. The current on a cell edge is then simply the charge flowing through the associated cell face over the course of a time step.

The problem with this straightforward procedure is that the current deposited in this way does not satisfy the discrete continuity equation given by

$$\begin{aligned} & \frac{\rho(i\Delta x, j\Delta y, t + \Delta t) - \rho(i\Delta x, j\Delta y, t)}{\Delta t} \\ & + \frac{j_x((i + 1/2)\Delta x, j\Delta y, t + \Delta t/2) - j_x((i - 1/2)\Delta x, j\Delta y, t + \Delta t/2)}{\Delta x} \\ & + \frac{j_y(i\Delta x, (j + 1/2)\Delta y, t + \Delta t/2) - j_y(i\Delta x, (j - 1/2)\Delta y, t + \Delta t/2)}{\Delta y} = 0 \end{aligned} \quad (6.42)$$

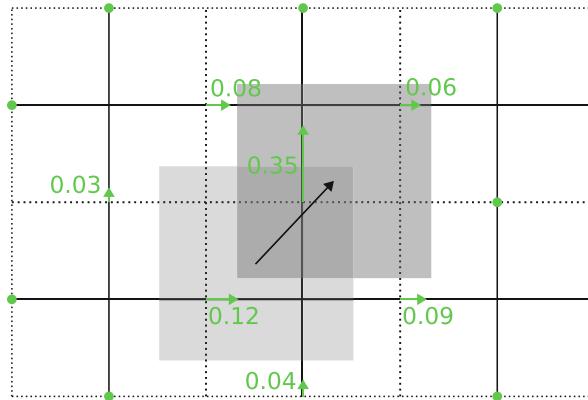


Fig. 6.6 Naïve current deposition for a charged particle moving as indicated by the black arrow and deposition current density onto the grid as shown in green. Solid black lines indicate grid cells, dashed line illustrate midplanes of cell faces

Since this current enters Maxwell's equations as a source term, it will lead to an electric field that does not actually satisfy Gauss's law. In other words the curl-free component of the electric field in an electromagnetic code using this current deposition scheme will differ from the electric field in an electrostatic code with the same particle distribution. This is clearly not desirable, and different methods have been used to correct the problem. One is to deposit the charge density separately and to solve a Poisson equation to get the longitudinal part of the electric field correct. It is computationally cheaper and very effective to make use of the fact that the electric field is nearly correct and to clean up the remaining error using a correction scheme, such as the one by Marder [47].

This problem is solved altogether by current deposition schemes that deposit a current that satisfies the discrete continuity equation directly. The first of these so-called *charge-conserving* schemes is by Eastwood [6], but several others have been developed since. The deposition in these schemes is slightly more expensive than in the naive algorithm, but the need to remove violations of Gauss's law is completely eliminated.

Somewhat more widely implemented in the scheme by Villasenor and Buneman [48]. The downside of this scheme is the need to treat the particle motion differently depending on how many cell interfaces the particle interacts with. Since this scheme needs to branch to different algorithms, its use on GPU is disadvantageous as branching is computationally more expensive there. Additionally this scheme is tedious to extend to shape functions beyond CIC.

A scheme that works with arbitrary shape functions (as long as they are products of separate one-dimensional shape functions) and without the need for different treatment of special cases is the current deposition by Esirkepov [49]. Figure 6.7(left) illustrates the current density that is deposited using this algorithm for the same particle move as used in Fig. 6.6. The reader can verify that the total

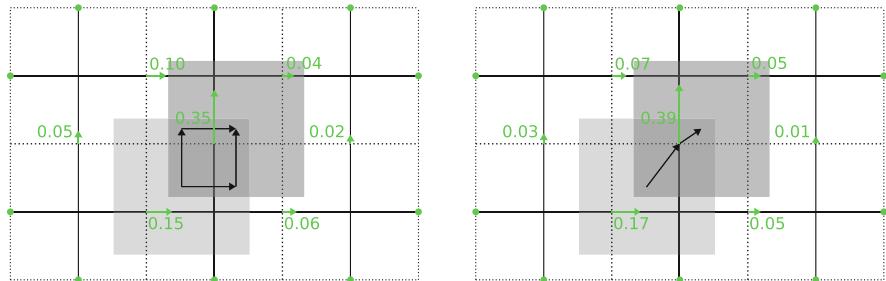


Fig. 6.7 Two options for charge-conserving current deposition. *Left:* Current deposition using the algorithm by Esirkepov [49]. The particle motion is decomposed into axis-parallel segments that are individually weighted. *Right:* Current deposition using the Zig-Zag algorithm by Umeda et al. [50]. As the particle closes from one cell to another, its trajectory is modified to go via the midpoint

current is identical, but the distribution onto the different grid points is different. In this algorithm the motion of the particle is decomposed into separate displacements parallel to the grid axis that are weighted individually. Readers that are interested in implementing this algorithm are referred back to the excellent original paper.

The other widely used scheme is the Zig-Zag scheme described in [50]. In this scheme particles can only cross between cells at the midpoint of the interface. A particle that does not naturally go through that point has its motion split into a motion from the starting position to the midpoint, followed by a separate motion from the midpoint to the final location, resulting in a Zig-Zag trajectory as illustrated in Fig. 6.7 (right). Although the method requires some code depending on conditional branching, the computational cost and noise properties between the latter two deposition schemes are similar.

Once the current is deposited on the grid, it can be filtered to remove unwanted noise. The noise level in PiC simulations is exacerbated by the artificially lowered number of macroparticles. The paper by Vay et al. [24] has an introduction to some digital filters and the associated terminology. Many codes simply use one or more passes of a simple binomial filter to reduce high-frequency fluctuations, but it is possible to tune filters specifically to the problem at hand and get much better performance.

6.2.5 Initialization and Particle Handling

So far we have described how the temporal evolution of particles and fields and their interaction are done in PiC codes. Yet the initial conditions also have to be prescribed. For field quantities this is a straightforward procedure: Analytic expressions for equilibria are used (typically the magnetic field is given, and the electric is assumed to be zero). Reproducing the desired distribution function $f(\mathbf{x}, \mathbf{v}, t_0)$ is more complicated, and the method used may have a measurable impact on the noise level.

The easiest and not uncommon case is a constant spatial density and a non-relativistic Maxwellian velocity distribution. In this case we can pick the initial position of the particle at random using

$$\mathbf{x} = (L_x \cdot \mathcal{U}(0, 1), L_y \cdot \mathcal{U}(0, 1), L_z \cdot \mathcal{U}(0, 1)) \quad (6.43)$$

where the random generator function \mathcal{U} produces three independent (pseudo-) random numbers, uniformly distributed between zero and one. The expected number of *particles per cell (PPC)* is Poisson distributed as we have a discrete probability distribution with independent probability. Since the number of macroparticles is significantly smaller than the number of real particles, the statistical density fluctuations are higher in the simulation than in the reality. One easy step to remedy this to some extent is to create exactly the expected number of macroparticles per cell and only choose the position inside the cell at random. This eliminates the fluctuations in the number of particles per cell and also makes it easy to represent spatially changing densities, by computing the number of particles in each cell from a known analytic function $n(\mathbf{x})$. This works well as long as the scale length $l_n = n / \frac{\partial n}{\partial x}$ is large compared to the cell size Δx .

For uniform densities there is also the possibility to use *low-discrepancy series* such as the Halton sequence [51]. This sequence distributes particles more uniformly in the simulation domain. This reduces the noise level as $\mathcal{O}(1/N)$ with the particle number. Adding more particles at random would reduce the noise level only like $\mathcal{O}(1/\sqrt{N})$. This *quiet start* has been successfully used, e.g., in the study of instabilities (see [52] for an example). But beware the possibility of multi-beam and ring instabilities as analyzed by Birdsall and Langdon [17].

The non-relativistic Maxwellian distribution is simply the composition of three velocity components that individually follow a normal distribution:

$$\mathbf{v} = (\mathcal{N}(u_x, v_{\text{th}}), \mathcal{N}(u_y, v_{\text{th}}), \mathcal{N}(u_z, v_{\text{th}})) \quad (6.44)$$

$\mathcal{N}(\mu, \sigma)$ produces a random number from a normal distribution with mean μ and standard deviation σ . The mean can be the average drift velocity of that species (and can depend on position and particle species, e.g., in the setup of a current sheet), and the standard deviation is set by the thermal speed

$$v_{\text{th}} = \sqrt{\frac{k_B T}{m}} \quad (6.45)$$

Note that there is a second popular definition of the thermal speed that includes a factor of square root two. When comparing with other codes and results, this is a point worth double-checking. Different species of different masses will have different thermal speeds at the same temperature.

If the drift speed \mathbf{u} approaches relativistic speeds, but the thermal speeds are still sufficiently low, some accuracy can be gained by generating a non-drifting distribution and boosting it with a Lorentz boost. If the thermal speed is large as

well, it is necessary to generate a proper Maxwell-Jüttner distribution. This used to be problematic, and many older codes neglect this; however this leads to problematic deviations from the correct distribution as demonstrated in [53]. Fortunately an efficient algorithm is described in [54].

Sometimes it is desired to generate new particles from a given distribution during initialization, but also later in the simulation, e.g., to fill a portion of newly expanded simulation space or to inject plasma through an inflow boundary. This is possible, but some subtleties need to be considered due to the interaction with existing fields in the domain [55].

In many simulations particles are created once, and afterward their number never changes. It can however be useful to split particles into particles with smaller macrofactor in regions of (phase-)space, where there is a low number of particles, to get better statistics and reduce noise. Many of those particles however leave the interesting region again and need to be merged together again to avoid a ballooning of the computational cost.

Splitting particles in a way that conserves mass, momentum, and energy is straightforward, as we can simply pick a source particle, split it into two equal parts of half the statistical weight and displace them slightly from the center mass to ensure that the following plasma dynamics will transport them separately. Some care can be taken that the displacement does not change the energy in the electromagnetic fields. The main problem is particle merging, as we will never find two particles of exactly equal velocities close to each other. Combining two particles of unequal speeds however cannot conserve both momentum and energy at the same time. In simulations where many new particles are created, the algorithm by Welch et al. [56] that estimates the distribution function from $N \gg 2$ particles and replaces them with M ($2 \ll M < N$) new particles can be useful. The method of [57] extends this to include also the stress tensor. If we do not want to redraw the complete particle population, e.g., because we want to preserve features in phase space that are not well described by the first few fluid moments, it is necessary to find particles of similar speeds. The method by Martin and Cambier [58] to find such particle pairs or groups can be useful for this task. The method by Vranic et al. [59] aims to merge N existing particles into two new particles while conserving charge, momentum, and energy.

6.2.6 Boundary Conditions

As mentioned before it might be desirable to create new particles at the boundary, and proper care has to be taken to match the existing fields there. This is a general problem with boundary conditions in particle-in-cell codes. One has to handle both the fields and the particles there and should do so in a consistent way to avoid problems.

Numerically easiest are periodic boundary conditions where the cell at the upper end of a coordinate is logically wrapped around (for N cells cell number 0 is right of cell $N - 1$) to be just below the lower end of the coordinate. Particles that leave

through one side are reinjected unmodified at the other end. Periodic boundary conditions mimic infinite space reasonably well, as there are no fixed walls or ends to the simulations. The periodic nature of the simulation does however select a finite number of modes that have an integer number of wavelengths in the simulation box. The number of grid points limits the mode number and the spectrum.

Also popular are Dirichlet boundary conditions for some field components: The normal component of the magnetic field is held fixed in time, and the two tangential components of the electric field are set to zero. Geometrically this affects the three field components that are located at that face of the unit cell. The two magnetic field components on the perpendicular faces and the electric field on the edge normal to the boundary are not needed at the boundary and are evolved freely, same as in the interior of the domain. Vanishing electric field, even in the presence of an electric current, is typical for a *perfect electric conductor* (PEC). This boundary therefore mimics a metal wall with vanishing resistivity. If a realistic metal boundary is required, much more effort is needed. This boundary condition does reflect impinging electromagnetic waves. The PEC boundary is therefore often coupled with a reflecting boundary condition for particles. When a particle ends outside the boundary at the end of the time step, both the component of the position normal to the wall and the normal component of the velocity are flipped. This process is identical to specular reflection when reaching the boundary.

More complicated boundary conditions for particles might absorb the particle or emit one or more new particles drawn from some distribution function, possibly depending on the energy of the incoming particle. This can simulate more or less realistic metal wall including effects such as wall charging or secondary electron emission.

While physically not a boundary condition, moving simulation domains are often implemented in terms of boundary conditions. One example is the moving window in the XOPIC code [60]. Here the Eulerian mesh is not moving, but instead new grid cells are created on one side and populated with background magnetic field and fresh thermal particles. On the other side, the grid is truncated and particles are removed. If the window happens to move at the speed of light—common in simulation that follow the path of a laser pulse through a plasma (see, e.g., [61])—no perturbation caused by the downstream boundary can reach the interior of the simulation domain, greatly simplifying the implementation. If the window is moving slower, a boundary that absorbs waves and does not reflect them back into the interior is needed at the lower end of the moving window.

This brings us to the topic of absorbing boundary conditions. These have been a problem for a long time. Initially Liao boundary conditions [62] and Mur [63] have been used. These boundary condition are described in detail in the great book on FDTD methods by Taflove and Hagness [64]. The problem with these boundary conditions is that they rely on fixed wave speeds. The presence of the plasma changes the wave speed. This makes it very hard to absorb all wave modes that reach the boundary sufficiently well.

A great step forward was the development of the *perfectly matched layer* (PML) by Berenger [65]. This layer is technically not a boundary condition, but rather

a region of space with modified permeability and permittivity. These material properties are chosen such that the impedance of the material is identical to the adjacent free space. Therefore the refractive index does not change from free space to the material layer, and no waves are reflected. Inside the material the wave is however damped, attenuating before it reaches the actual boundary. There a simple boundary condition such as PEC can be chosen. Any wave that does reach the boundary and gets reflected is attenuated a second time on the way out, greatly improving performance of this scheme. One implementation known as convolutional PML (CPML) has been used successfully in multiple cases. The book [66] explains the different methods and their connections.

PML boundary conditions are often combined with boundaries that absorb outgoing particles and replace them in some scheme to simulate the presence of a thermal plasma bath outside the simulation domain that can act as the source of unperturbed upstream. This has been used in the simulation of magnetic reconnection, where the removal of plasma in the outflow regions and the injection of fresh upstream with not-yet-reconnected magnetic flux allows for the study of larger systems (see [4, 67, 68]).

It is also possible to inject electromagnetic waves to study their propagation. This can either be achieved by the means of an artificial current, acting as an antenna, or by manipulating the electromagnetic field directly. More complicated injection schemes that allow for the injection of realistic laser pulses [69, 70], of oblique waves with arbitrary envelops [71] or of magnetized plasma without unwanted numerical noise [72] are possible as well.

6.2.7 Parallelization

Hopefully you are convinced by now that particle-in-cell codes are actually a fairly approachable simulation method that can give great insights into the microphysics of kinetic plasmas. But there is one downside compared to other methods explained in this book: A huge number of particles and a large number of time steps are required to simulate physically relevant systems. And even then we are not covering macroscopic scales. To increase the maximum size of the problem that can be studied and to reduce the time a scientist who has to wait before obtaining results that can be studied, parallelization is extremely common and has been used for a long time.

This parallelization can make use of vectorization in the form of single instruction, multiple data (SIMD) [11]. Details of this depend on the platform that the code targets (CPUs, GPUs, accelerator cards, and so on). The recent arrival of performance portability frameworks such as Kokkos [73] or RAJA [74] is changing the codes from carefully formulated C++ and Fortran codes that can be vectorized by the compiler and handwritten assembler primitives to higher levels of abstraction. While good reference literature on the topic is rare, the field is changing too rapidly for us to give good advice that will still be relevant in one or two decades.

The next level of parallelization that is often employed is the use of parallel threads within a process. This might be OpenMP threads [75, 76], pthreads, or other language-level features for parallelism. This can be used for the multiple cores that all current computers offer. Often the user has to manually ensure that no data races or other problems between different threads exist. Since this can be hard to ensure, some codes opt not to make use of these techniques.

Since the compute power of a single machine is only sufficient for very selected problems and the memory can only hold approximately one billion particles in RAM, it is commonly required to parallelize the code over many nodes. Each node runs one (or more) separate process that communicate with each other and share the simulation workload. For the communication MPI [77] is the de facto standard, but recently frameworks [78] have started to appear that take care of threading and communication for the programmer.

The common way to distribute the workload is to split the simulation domain into smaller local domains (also known as patches or blocks). Each process is then responsible for the particles and fields in that local domain [79–83]. The state of the electromagnetic field in neighboring domains is then copied into ghost zones that surround the local domain and act as boundary conditions.² Particles that hit the boundary of the local domain are communicated to the neighboring process, inject from the wall there, and are removed from the local domain. Particle-in-cell codes tend to parallelize very efficiently using this method since only communication with a handful of neighbors is required and global all-to-all messaging can be nearly completely avoided. The only caveat is that efficiency only remains high as long as all processes have fairly balanced workloads. If the setup starts unbalanced (e.g., because of a bad decomposition or because of inhomogeneous plasma density) or develops load imbalances over the course of the simulation (typically in reconnection simulations when plasmoids are generated or shock simulations where the downstream has higher density than the upstream), the parallelization efficiency drops. One can attempt to improve load balance in the simulation by moving the boundaries between local domains and making additional grid cells and particles the responsibility of less loaded processes, but this requires much more complicated code and incurs additional costs.

An alternative to splitting the global domain into rectangular or brick-shaped local domains is to compute a space-filling curve that forms a 1D path that goes through all grid cells exactly once [84–87]. Additionally the space-filling curve has the properties that grid cells that are close to each other along the curve are also mostly close to each other in the full 2D or 3D space. When the curve is then cut into segments to be distributed across processes, the cells of each process tend to be close to each other. For most of its grid cells, the process also owns the neighboring field values and can use those without additional communication. Moving particles that

² The reason why this works efficiently is that the particle shape function have only compact support and the field update only requiring neighboring cells. However with extra effort domain decomposition can also be applied to spectral solvers[27].

enter another grid cell are likely to stay with the same process. If they do leave, they are communicated to the correct neighboring process same as in the case with patch-based domain decomposition. The big upside is that the segments of the space-filling curve do not have to be equally long and can be rebalanced more easily.

6.2.8 Diagnostics

Designing and using a particle-in-cell code are interesting by itself, but the ultimate goal is to produce output that can be used to gain understanding of the physical system. This requires the calculation of diagnostic quantities and storing them in a way that persists beyond the duration of the simulation. Simply storing the full state of the simulation is typically impossible for performance reasons. Whenever possible the simulation code should calculate diagnostics as close to the final plots as possible and output only them. After all, a simple histogram for a velocity distribution might only be a few kilobytes, where the dump of all particles is several terabytes. Saving the simulation state for later analysis or restart might still be necessary but should be performed as rarely as possible.

There are of course many high-level diagnostics, and it is impossible to cover all of them. This section will therefore mostly cover general-purpose output that can be post-processed further. To gain high-level insight into the simulation at low I/O cost, it is often useful to calculate a breakdown of the energy content in the simulation in several rough categories, such as the energy in the electric field and the energy in the magnetic field and in the particles. The diagnostic is also very valuable to check conservation of total energy in a PiC code that conserves momentum by construction.

Another obvious candidate for output are gridded quantities, such as electric and magnetic field as well as current density. These quantities do not require any extra calculation. The two most widely used choices are dumping the memory of every local domain (in a simulation that uses domain decomposition and MPI) to a separate binary file or writing a global view to a structured format such as HDF5.³ The first choice has the advantage of low overhead as nearly no metadata is saved. The downside is that the separate views of the different local domains need to be combined for further analysis. The overhead of creating many files at once and the tedious file handling make this alternative a poor choice for more than a few hundred cores. Parallel I/O to HDF5 requires slightly more complicated I/O code and introduces some overhead but allows for the inclusion of metadata and easy loading into analysis tool, independent of programming language and CPU architecture.

Other field data that need to be computed just for diagnostic purposes such as the mass density, charge density, average bulk velocities, or components of the pressure tensor can also be good choices for grid-based output. Even if many of

³ Other formats are available, but HDF5 is widely supported by many post-processing tools.

these quantities are required, it is typically cheaper to compute the diagnostics during the parallel simulation instead of dumping the full particle information and doing post-processing.

In many cases not all of the three-dimensional field data is required. Averaging over one or two dimensions to obtain line profiles or two-dimensional selections is an option. Alternatively line-outs or two-dimensional cuts can be useful. If points of interest are known beforehand or can easily be computed on the fly, it is interesting to output time histories of those points at high cadence. If a global overview is needed, but not the full resolutions, it is possible to output a downsampled grid. Care should be taken to apply a digital low-pass filter—such as a Kaiser window [88]—before the decimation to avoid aliasing.

In some cases it is necessary to restart the simulation, either due to reaching a wall-clock limit or due to a compute node crashing. Also some analysis can only be performed as post-processing, either due to poor scalability or because it is actively being developed based on the simulation results. In this case, the full particle information needs to be stored. Again direct binary dumps or higher-level self-describing formats are the two main possibilities.

Many other diagnostics have been developed, and the following list of examples is certainly incomplete: one-dimensional histograms of particle quantities (speed, energy), two-dimensional cuts in phase space showing, e.g., $x - v_x$ -densities, Fourier transforms of other quantities such as $B(|k|)$, time averages of other quantities to remove high-frequency noise, trajectories of specially selected particles of interest, often together with local field quantities at the position of the tracer particles, reconstruction of the particle phase space density on a six-dimensional grid, and discrete estimates of the entropy based on a binned phase space. The possibilities are endless here, and a large fraction of the development time of a mature code will be invested in providing user-friendly, scalable, efficient diagnostics.

One special diagnostic that is of interest in astrophysical applications is the calculation of radiation spectra. This is different from the spectrum of the field inside the simulation domain that is limited to frequencies that are explicitly resolved. Often there is synchrotron or inverse-Compton radiation at much higher frequencies that can be calculated separately. See [89] on how to implement and properly test such radiation diagnostics. Some recent codes include the production of gamma rays and electron-positron pairs [90]. This is especially relevant for the nascent field of pulsar magnetosphere simulations, where the back reaction of the radiation processes on the energetic particles in the simulation needs to be included [91].

6.2.9 Test Cases

Compared to fluid codes that easily develop instabilities, e.g., through negative densities or numerical instabilities, PiC codes are rather robust. This numerical stability is sometimes mistaken as physical plausibility. It is therefore important that the researcher employs a healthy skepticism toward the code and that the code is tested. Other communities like the fluid simulation community have a large number

of test cases for which analytic solutions are known or at least results from other codes are available. In the particle-in-cell community validation efforts (making sure that the employed plasma model is adequate and that the underlying assumptions are met), benchmarking (comparison with other codes) and validation (testing proper implementation) are relatively new and rare. One of the reasons is that the number of test problems with known, analytic solutions is rather small.

But there are some test cases that can be recommended. The paper by Jacobs and Hesthaven [92] introduced a discontinuous Galerkin code on unstructured mesh but also includes several useful test cases. The study of Kelvin-Helmholtz instability in Saturn's magnetopause [93] has been used by several hybrid codes. Code description paper such as the one for the Apart-T code [53] or the SHARP code [14] lists several test problems that are solvable with that code and offers opportunity for benchmarking. Papers that investigate new hybrid codes such as [94, 95] contain test problems to demonstrate the new method, but many of them are also accessible to PiC codes. The paper by Kilian et al. [96] contains a list of wave modes and suggested parameters that are accessible to different plasma models that can be implemented by PiC codes.

Physical setups for which analytic solutions exist include the study of plasma waves [97–100] and the thermal excitation of fluctuations in density [101] or electromagnetic fields [102, 103]. For a complicated distribution function that does not admit analytic solutions for the dispersion relation of plasma waves, comparison with numeric dispersion solver [104, 105] is useful. The study of Landau damping has a long tradition [106] and can be useful to test the long-term stability of the code [107].

The comparison between different code types such as in [108, 109] can be useful to find the limits of applicability. This kind of comparison is especially popular in the fusion power community [110]. The expected scaling behavior when improving numerical resolution can also be used to test the code, even without outside reference simulations, as shown in [111].

6.3 Applications of Particle-in-Cell Simulations

6.3.1 Turbulence in the Kinetic Regime

There is little doubt that kinetic plasma can be turbulent, but theories to describe such turbulence are so scarce that simulations of turbulent kinetic plasmas will be needed for many decades. One reason for the lack of theories is that already MHD turbulence is little understood. The Kolmogorov theory proposed energy transfers from large eddies to small eddies, and this is not doubted—although many questions are still open.

The attempt of Kraichnan and Iroshnikov to transfer the ideas from hydrodynamics to magnetohydrodynamics required the use of plasma waves instead of eddies. The question whether plasma turbulence can be described as wave-like turbulence

is still disputed. The theory of Goldreich and Sridhar which addresses the question of anisotropy in magnetized plasmas is also based on the wave approach.

All major theories for plasma turbulence are limited to MHD plasma more specifically to incompressible MHD plasmas. The main reason is the simplicity of the dispersion relation of waves in this case: Only Alfvén waves with $\omega^2 = k_{\parallel}^2 v_A^2$ exist, which makes analytical calculations feasible in the first place.

The regime of a linear dispersion relation is limited mostly to frequencies well below the ion-gyrofrequency. Beyond this point dispersive effects play a role. In the region close to the ion-gyrofrequency, various approximations can be made like the electron-MHD [112] or the Hall-MHD [113] approximation.

6.3.1.1 Initializing Turbulence

Turbulence models typically assume an energy cascade from large scales to small scales. In order to initialize a turbulence simulation, fluctuations have to be injected at large scales. There is an extensive discussion about the nature of turbulence—waves or structures [114]—but we will focus on the initialization via waves here.

There are two types of turbulence simulations commonly used: driven and decaying turbulence. In the latter case, waves or structures are initialized at the start of the simulation, and no further energy is injected. The former case has energy injection—typically waves at low k —continuously. While there are physical scenarios for both cases, only the continuous injection may lead to an equilibrium.

There is a large variety of plasma waves in the kinetic regime. Their dispersion relation can typically be derived by solving for small fluctuations in the dielectric tensor. Self-consistence then requires that the distribution in position and velocity of the background plasma particles has to be adjusted as well.

Schreiner et al. [115] have performed the necessary calculations for L- and R-modes (in the subluminal branch). The necessary initial conditions for L-waves (R-waves are analogous) are derived from the dispersion relation

$$k_L = \pm \frac{\omega}{c} \sqrt{1 - \frac{\omega_p^2}{(\omega + \Omega_e)(\omega - \Omega_p)}} \quad (6.46)$$

The wave number k_0 for the excited wave has to be chosen as an integer multiple of $2\pi/L$. Now Eq. (6.46) has to be solved for the frequency ω_0 associated with k_0 . The amplitude of the wave is defined by the absolute of the electric field fluctuation δE :

$$\delta E_y = -\delta E \sin(k_0 x), \quad \delta E_z = -\delta E \cos(k_0 x), \quad (6.47a)$$

$$\delta B_y = -\frac{c k_0}{\omega_0} \delta E_z, \quad \delta B_z = \frac{c k_0}{\omega_0} \delta E_y \quad (6.47b)$$

The magnetic field fluctuation is calculated directly from Maxwell's equations. The choice of the left-handed L-wave determines the polarization and in turn the relation of the y and z component of electric and magnetic field.

At the start of the simulation, the electromagnetic fields are stored in the cells according to Eq. (6.47). The initially thermal distribution of ions and electrons has to be boosted to match the currents associated with the transverse wave. The boost velocity δv for the L-wave can be calculated as well:

$$\delta v_y = \frac{q_\alpha c \delta E_z}{q_\alpha B_{\text{ext}} - c m_\alpha \omega_0}, \quad \delta v_z = \frac{-q_\alpha c \delta E_y}{q_\alpha B_{\text{ext}} - c m_\alpha \omega_0} \quad (6.48)$$

The boost velocities for species α , electrons or protons, depend on their charge q_α and mass m_α .

6.3.1.2 Analyzing Turbulence

Following Kolmogorov's idea the most important quantity in the analysis of turbulence is the one-dimensional Fourier energy spectrum. This clearly shows how energy is cascading (typically from low to high k). The 1D energy spectrum can prove the existence of an energy cascade, but in magnetized plasmas with a possible anisotropy, this tool is not helpful. Following the Goldreich and Sridhar theory for magnetized incompressible MHD plasmas, the anisotropy should develop like this: There is generally only transport of energy perpendicular to the magnetic field (k_\perp) until at some point the so-called critical balance occurs, where transport along k_\parallel is possible through nonlinear effects.

While kinetic turbulence only remotely resembles incompressible MHD turbulence, there are still waves with similar dispersion relations (R-modes in the first place) that could lead to a similar behavior. This can be seen very well in the two-dimensional (k_\parallel, k_\perp) energy spectrum. Real space plots of turbulent media are usually not helpful unless structures (e.g., islands or current sheets) emerge that can be seen in real space.

For driven turbulence one hopes to find an equilibrium state which makes further analysis easier, but the continuous injection of energy will heat the plasma and increases the thermal energy of the particles. Even the fluctuations in the electromagnetic field might never be stationary, so the determination of initial transient behavior becomes somewhat subjective.

6.3.1.3 Exemplary Results

Simulations of kinetic turbulence differ in a large number of input parameters. Compared to hydrodynamic turbulence, there are additional magnetic field parameters, and compared to MHD simulations, an absolute outer length scale is added. PiC simulations may add numerical parameters like the electron-to-proton mass ratio. The parameter space is therefore too large to give a complete overview.

Here we will show only few sketches to highlight how a possible setup of turbulence simulations looks like and what diagnostics are possible. The reader is referred to some recent publications in the field including a comparison of hybrid

and PiC simulations of turbulence [116], 3D simulations of turbulence with the OSIRIS code [117], more 3D simulations [118], and simulations of turbulence in a pair plasma [119].

The physical and numerical parameters of the simulation presented here are given in Tables 6.1 and 6.2, respectively. A central problem in finding a suitable parameter set is the balance between the requirement to resolve the Debye length and the onset of cyclotron damping. Energy will not be transferred to large parts of k space due to damping processes.

The simulation is initialized with waves at small k around $k = 0$. The distribution of energy in the two-dimensional k space is then depicted in Fig. 6.8. Here a picture very similar to the Goldreich-Sridhar theory emerges: Energy is transferred primarily perpendicular to the magnetic field (further simulations not shown here, show that even for initial energy distributions elongated parallel to the magnetic field show a similar picture).

Table 6.1 Physical parameters for the turbulence simulation: plasma frequency $\omega_{p,e}$, cyclotron frequency Ω_e , and thermal speed $v_{th,e}$ of the electrons, the magnetic fluctuation strength δB^2 , and the plasma beta β

$\omega_{p,e}$ (rad s $^{-1}$)	$ \Omega_e (\omega_{p,e})$	$v_{th,e}$ (c)	$\delta B^2/B_0^2$	β
$1.966 \cdot 10^8$	0.447	0.10	0.10	0.20

Table 6.2 Numerical parameters for the 2D simulation: number of cells N_{\parallel} and N_{\perp} in the directions parallel/perpendicular to the background magnetic field \mathbf{B}_0 , number of time steps N_t , grid spacing Δx , time step length Δt , and PPC

$N_{\parallel} (\Delta x)$	$N_{\perp} (\Delta x)$	$N_t (\Delta t)$	$\Delta x (c \omega_{p,e}^{-1})$	$\Delta t (\omega_{p,e}^{-1})$	PPC
1024	1024	$1.0 \cdot 10^5$	0.10	0.05	128

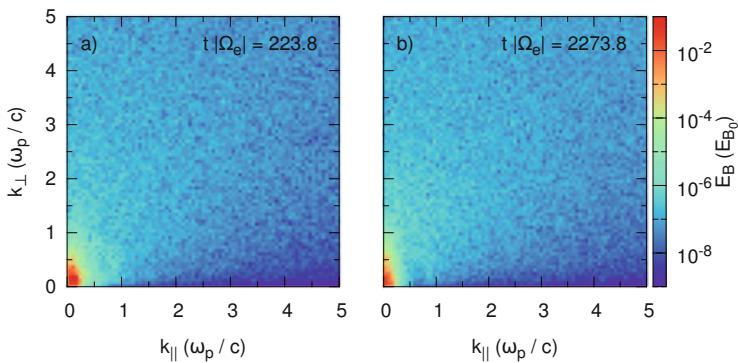


Fig. 6.8 Magnetic energy spectrum at two points in time. The energy density is normalized to the background magnetic field

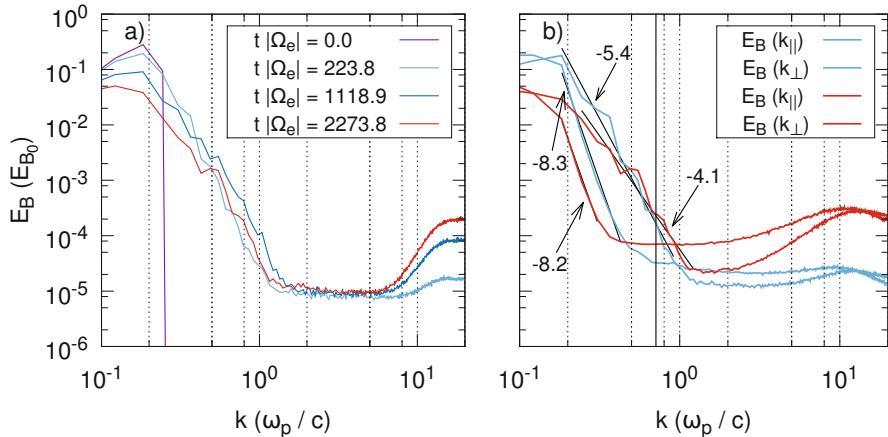


Fig. 6.9 One-dimensional energy spectra normalized to the background magnetic field. Both panels depict the magnetic energy spectrum at different times, panel (a) shows the magnetic energy spectrum at different times, while panel (b) shows parallel and perpendicular spectrum separately. The time steps are the same as in simulation as the data in Fig. 6.8. The vertical black line near $k c/\omega_{pe} \approx 0.7$ marks the onset of cyclotron damping of parallel propagating waves

The 1D Fourier spectra parallel and perpendicular to the magnetic field shown in Fig. 6.9 provided different pictures. The general spectrum is much steeper than expected from classic theories.

As discussed before these simulations can only give insight into a small part of the very large parameter space of kinetic turbulence. While more simulations are required, the computational cost for any interesting physical parameter set is mostly extreme. Only time and more computer power can give more insight.

6.3.2 Charged Particle Transport

High-energy astrophysics does not only cover the acceleration of particles to the highest energies at their source, but includes also the transport of energetic particles to the observer. Observations of energetic particles can in fact only be understood, when their transport is understood.

Although the plasma background and the energy of particles differ drastically between intergalactic, interstellar, and interplanetary media, the governing physics behind all three scenarios is kinetic plasma physics. This does not only mean that the spectrum of particles is non-Maxwellian, but it can also mean that the plasma background includes dispersive waves.

The theory of wave-particle interaction and the resulting transport parameters have been discussed extensively in the past three decades, and the reader is directed to the text book by Schlickeiser [120] for a detailed introduction. The quintessence of theoretical efforts is that any analytical calculation is limited to simple cases.

Numerical models have evolved over time to fill this gap: starting with simple test particle models with single wave modes to models with artificial turbulent spectra leading to full MHD models [121–123].

Using PiC it is possible to assume non-test particles in kinetic turbulence. As it will be shown, this approach is allowed within the PiC framework, but not (yet) feasible. To study the interaction of particles with a single wave, these waves have to be initialized in the first place. The process of wave creation has been discussed in detail in Sect. 6.3.1.1. Which waves or more specifically which wave spectrum is going to be initialized depends on the scientific question:

- Single waves can be used to compare analytic results with numerical experiments. Some benchmark results will be shown below.
- A spectrum of waves with random phases can be used to study transport in artificial turbulence.
- Finally a turbulent spectrum as discussed in the previous section can be used as target field for particle transport.

In order to study the transport, a species of particles whose transport properties are to be examined will be injected. Typically suprothermal particles are used. Here care has to be taken, which physical effect should be studied:

- Test particles (or typically for the sake of simplicity: particles with very small macrofactor) can be used to study the passive transport in a kinetic plasma. This is also the approach used in MHD plasmas.
- Full particles (along with the correct physical distribution) can be used to study transport of particles in a plasma, where instabilities are excited by the particles. Here the considerations discussed in Sect. 6.3.3 have to be taken into account regarding the timescales and the necessary simulation size.

Regardless of the approach, the transported particles should be marked as a separate species for later analysis purposes.

In transport studies typically three quantities are scientifically interesting the pitch-angle diffusion coefficient $D_{\mu\mu}$, the momentum diffusion coefficient D_{pp} , and the perpendicular spatial diffusion coefficient D_{\perp} . From the definition

$$D_{\mu\mu} = \lim_{t \rightarrow \infty} \frac{(\Delta\mu)^2}{2\Delta t} \quad (6.49)$$

and similar expressions for D_{pp} and D_{\perp} , the transport coefficient can immediately be deduced by calculating $\Delta\mu = \mu(t) - \mu_0$ individually for every particle for long enough times t . The transport coefficient can then be binned in pitch-angle μ and/or momentum p according to the scientific question. It should be noted however that this method called “running diffusion coefficient” is error prone—especially for the pitch angle and for strong turbulence: Only for very small change in the pitch-angle

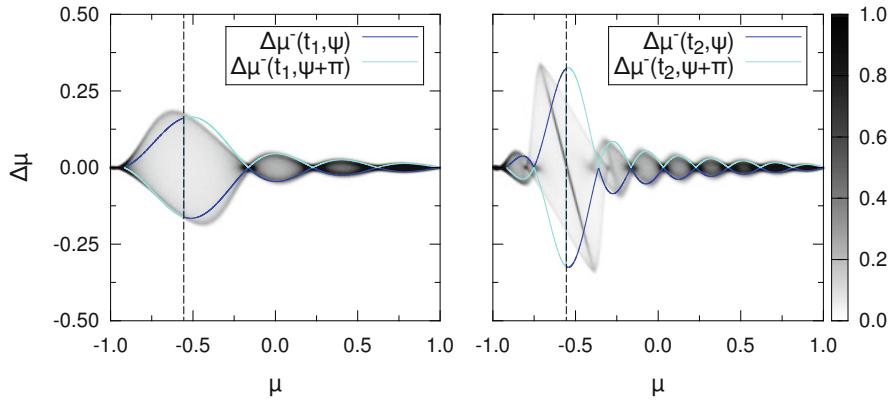


Fig. 6.10 Pitch-angle scatter plots present the change $\Delta\mu$ of a particle's pitch angle at time t depending on its initial μ at $t_0 = 0$, where t_1 and t_2 are intervals of two and four gyroperiods, respectively. Color coding indicates the number of particles at a given point $(\mu, \Delta\mu)$ normalized to the total number of test protons. The dashed lines show the position of the resonant pitch angle μ_{res} given by Eq. (6.50). The curves represent quasi-linear theory (QLT) predictions for the maximum scattering amplitude according to Eq. (6.51). Figure from [115] reprinted with permissions by ©Elsevier

$D_{\mu\mu}$ can be calculated as a function of μ . The reader is referred to [123] for an advanced analysis method based on the change of particle distribution functions.

Figure 6.10 shows the transport of energetic protons with velocity v_T and gyrofrequency Ω_T interacting with a single wave. The amplitude of the wave is between $0.021 < \delta B/B_0 < 0.048$. The expected behaviour is the resonance of particles with

$$k_{\parallel} v_{\parallel} - \omega_0 + n \Omega_T = 0 \quad (6.50)$$

The change of pitch-angle can be calculated

$$\begin{aligned} \Delta\mu^{\pm}(t, \psi) = & \Omega_T \sqrt{1 - \mu^2} \frac{\delta B}{B_{\text{ext}}} \times \\ & \frac{\cos \psi - \cos ((\pm k_{\parallel} v_T \mu - \Omega_T) t + \psi)}{\pm k_{\parallel} v_T \mu - \Omega_T}. \end{aligned} \quad (6.51)$$

The observable tilt between theory and numerical experiment is an artifact of defining the μ -axis: The tilt can be removed when choosing $1/2(\mu_{\text{end}} + \mu_0)$ on the μ -axis.

6.3.3 Instabilities

Instabilities play a crucial role in space plasmas. Many of the application are discussed throughout this book starting with MHD instabilities like Kelvin-Helmholtz in Chap. 1.4.1 (Introduction to MHD Simulations) and Kelvin-Helmholtz in Chap. 13.5.3 (EMAPS: An Intelligent Agent-Based Technology for Simulation of Multiscale Systems) to various kinetic instabilities (beam instability in Chap. 5.5.2 (Eulerian Approach to Solve the Vlasov Equation and Hybrid-Vlasov Simulations) and Chap. 10.7.2 (Generalized Quasi-Neutral Hybrid-Kinetic Simulations), instabilities in reconnection and shocks in Chap. 9.5.1.2 (Hybrid-Kinetic Approach: Inertial Electrons), and Bunemann, Weibel, and magnetorotational instability in Chap. 11.4 (Fully Kinetic (Particle-in-Cell) Simulation of Astrophysical Plasmas)).

PiC codes are able to reproduce kinetic instabilities that include but are not limited to cases with non-Maxwellian distribution functions, counterstreaming plasmas, or currents in plasmas. This section serves as a hands-on example on how to initialize, run, and evaluate PiC simulations of kinetic plasma instabilities.

PiC simulations of plasma instabilities date back to the early 1970s [124] where a prime example of a purely kinetic instability has been analyzed: the Weibel instability. As it turns out, a very large number of distinct instabilities can be found in kinetic plasmas. The fact that different instabilities can occur simultaneously in one plasma [125] is an important reason to use numerical experiments to study them. The complicated interplay of the instabilities and the late-time nonlinear behavior cannot be understood from analytical calculations. The reader is referred to [126] for a very detailed overview of instabilities (more specifically beam-plasma instabilities).

The classical approach to plasma instabilities is described in-depth in Chapter 11 of [120]: For a given configuration of the distribution function, the dielectric tensor of the plasma can be evaluated, and the dispersion relation for that plasma configuration can be solved. This typically yields solutions of the form

$$\omega(\mathbf{k}) = \omega_R(\mathbf{k}) + i\gamma(\mathbf{k}) \quad (6.52)$$

where ω_R is the real part of the frequency and γ is the growth rate. Three different cases can be distinguished:

- $\omega_R \neq 0, \gamma = 0$: Eigenmodes of the plasma that are not damped
- $\omega_R \neq 0, \gamma > 0$: Waves with frequency ω_R and growing amplitude
- $\omega_R = 0, \gamma > 0$: Non-propagating structure that grows in time

Analytic theory can answer the question how fast a given mode of the plasma is growing—should the equation for the dispersion relation be solvable. It should be noted that the evolution of the instability changes the plasma configuration. This leads to a nonlinear evolution, not described by analytical theory.

The first step in simulating instabilities is the initialization of the plasma configuration. Following the description in Sect. 6.2.5, particles will be injected.

For simple scenarios like counterstreaming plasmas with identical densities, this is a simple task. More involved cases are, e.g., cosmic-ray-induced instabilities with high-density contrasts. In order to accommodate high contrast, very high particle numbers can be used (which is discouraged by the computational cost), or the weight of macroparticles has to be changed.

Care has to be taken also for instabilities caused by spatial inhomogeneities: as the number of particles per cell at the point of lowest density should fulfill the minimal criteria for numerical stability (at least one particle per cell, typically 10 or more), which in turn requires again either higher particle numbers or different particle weights in regions with higher density.

A more important question is the correct resolution and size of the simulation. It is self-evident that the cell size has to resolve the Debye length in explicit PiC codes; on the other hand, the total size of the simulation has to be large enough to contain the waves or structures that are excited. This yields a minimal size $L_{\min} \geq 2\pi/k_{\max}$, where k_{\max} is the wave number, where the maximum growth of an instability is expected. It has been argued [127] that it is not sufficient to resolve k_{\max} but that it is necessary to resolve the width of the peak of the damping function $\gamma(k)$.

For most astrophysical plasmas, the necessary cell number from the above condition is typically very large. Combined with the fact that growth rates are often very low compared to the plasma frequency ω_{pe} , this yields typically rather long simulation times—usually beyond what is possible. In almost all simulations, other parameters are altered to produce computationally viable but still meaningful simulations. One of the most commonly used techniques is changing the mass ratio of protons and electrons. This reduces the time needed to see ion motion and therefore produces shorter simulations. Unfortunately not all physical effects scale in the same way with the electron-to-proton mass ratio. It is advisable to examine the dependence of growth rates on the mass ratio closely [128, 129]. As a rule of thumb, the change of mass ratio should not change the ordering of proton and electron gyro- and plasma frequencies.

Based on the output described in Sect. 6.2.8, there are various ways to find instabilities in the simulation data. The primary hint for the existence of an instability is the rise of particle or electromagnetic energy. In order to detect and distinguish different instabilities, it is inevitable to use Fourier transforms of the electromagnetic field data. The following recipe has proven itself to be very helpful:

1. Consider the transverse magnetic field as well as the curl-free longitudinal and divergence-free transverse components of the electric field.
2. Plot the energy density of these fields in the two-dimensional Fourier domain. If the simulation was done in 3D, plot both the plane of $k_{\parallel} - k_{\perp}$ and the plane with both directions perpendicular to the mean magnetic field.
3. The energy plot gives a hint at which points in time large changes happen in the simulation. For each of these times, the spectrum of fluctuations should be plotted and examined separately.

Following this recipe will provide an increase of energy as a function of the position in k -space. Together with the information whether this is a transverse or longitudinal instability, it should be possible to compare with theoretical predictions. Plotting field energy in, e.g., longitudinal and transverse electric field, at a given wave number \mathbf{k} over time yields the growth rate.

As the simulations resemble the realistic, nonlinear behavior of the instability, a comparison of growth rates with analytical theory is only possible for small times, where the instability is still in the linear regime.

An illustrative example is the instability caused by pair beams originating from TeV photons in the intergalactic medium. The scenario has been proposed by Neronov and Vovk [130] to explain the lack of GeV photons. While the physical problem at hand is only of special interest, it illustrates very well the difficulties in simulating (beam-driven) instabilities. The physical setup is such that electron-positron pairs with Lorentz factors of 10^{11} stream through hot gas. The density contrast between gas and beam is about 10^{13} . Translating these physical parameters into a numerical setup shows immediately that PiC simulations are not doable as is. The high-density contrast can only be achieved by altering the macrofactor for beam and background species, and the extreme Lorentz factor immediately leads to the numerical Cerenkov instability.

The simulations [131] and [132] show different approaches to adjusting physical parameters for a feasible simulation. While the simulations of Sironi et al. focus on achieving very high Lorentz factors for the beam, the simulations of Kempf et al. try to resemble the energy ratio of the beam to the background. Both simulations have in common that neither the real density contrast nor the Lorentz factor of the beam can be resembled.

In the further course of the simulations, two parameters have been defined: the mass density ratio α between a drifting species—denominated jet—and background, as well as the energy density ratio ϵ between jet and background. The physics case prompts for extremely small values of α and ϵ values well below 1. An interesting test case is the $\epsilon = 1$, $\gamma = 50$ case, because it illustrates rather well the onset of the change of the distribution function that is one source of nonlinear development of the instability.

In Fig. 6.11a the electric field energy plotted over time is used as primary marker for the instability. The time is normalized to inverse plasma timescales (a sensible approach, when the relevant times are not depending on the magnetic field). During the first 500 inverse plasma timescales, a sharp increase is seen, which is associated with the electrostatic instability. The growth rate decreases until around 3000 inverse plasma timescales, where a further rise associated presumably with the Weibel instability takes place (associated presumably with the Weibel instability).

The energy plot gives a hint, at which times a detailed analysis is helpful. When looking at times, where the saturation has set in, the results may be ambiguous. Here a look at the Fourier transform will yield further insight. The 1D (Fig. 6.11b) and 2D (Fig. 6.12a) Fourier transform is plotted. The 1D Fourier transform shows a clear sign of the instability at $4 \cdot 10^{-10} \text{ cm}^{-1}$ and another bump at $2.3 \cdot 10^{-10} \text{ cm}^{-1}$. The 2D Fourier transform shows that the former peak is linked to a structure at fixed

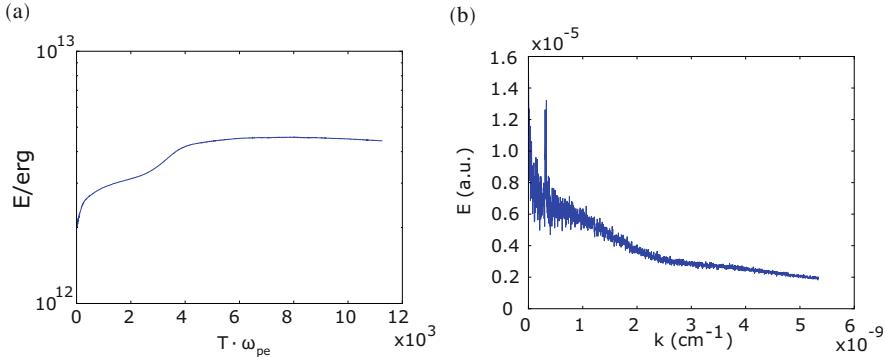


Fig. 6.11 Simulation of propagating pair beams with $\epsilon = 1$, $\alpha = 2.5 \cdot 10^{-5}$, $\gamma = 50$. (a) Electric field energy. (b) E-field 1D Fourier at $t = 11246\omega_{pe}^{-1}$

k_{\parallel} over a wide range of k_{\perp} . The 2D transform highlights that a 2D simulation is necessary to capture the features not aligned with the magnetic field direction. Here the position in k -space is compatible with $k_{\parallel}u \approx k_{\parallel}c = \omega_{pe}$.

When observing the change of the distribution function as outlined in Fig. 6.12b, it can be seen that the peak of the beam decreases in amplitude, but even after more than 10^4 plasma timescales, it just moves to a plateau. We conclude that the plateauing of the distribution function slows down the instability, which brings the system to an almost stable situation after around 7000 plasma timescales.

It is interesting to note that a clear nonlinear behavior can be observed even when the distribution function does not drastically change over time. The root cause for this is the possible nonlinear interaction of waves and structures excited by the instability itself. In the lowest order, this would be a three-wave interaction that can easily occur.

In a longer parameter study of this instability, it could be shown that the size of the simulation box has in fact an effect on the instability: It is a good practice to run such simulations at double resolution to see whether at least the energy gain converges.

While the latter example was from the class of beam-driven instabilities, there is another class of instabilities to which, for example, the filamentation and Weibel instability belong: instabilities which arise from counterstreaming plasmas or anisotropies in temperature. The simulations by Burkart et al. [128] will serve as an example for the filamentation instability in special.

The initialization of this family of instabilities is typically much simpler: As opposed to beam instabilities, there is usually not a large-density contrast. Our example for the filamentation instability assumes a proton-pair plasma running into a proton-electron plasma.

The setup is realized as two counterstreaming plasma populations, one representing the background medium consisting of $6e^-$ and $6p^+$ per cell (proton stream) and the other incorporating the jet containing $4e^-$, $2e^+$, and $2p^+$ per cell (pair-proton

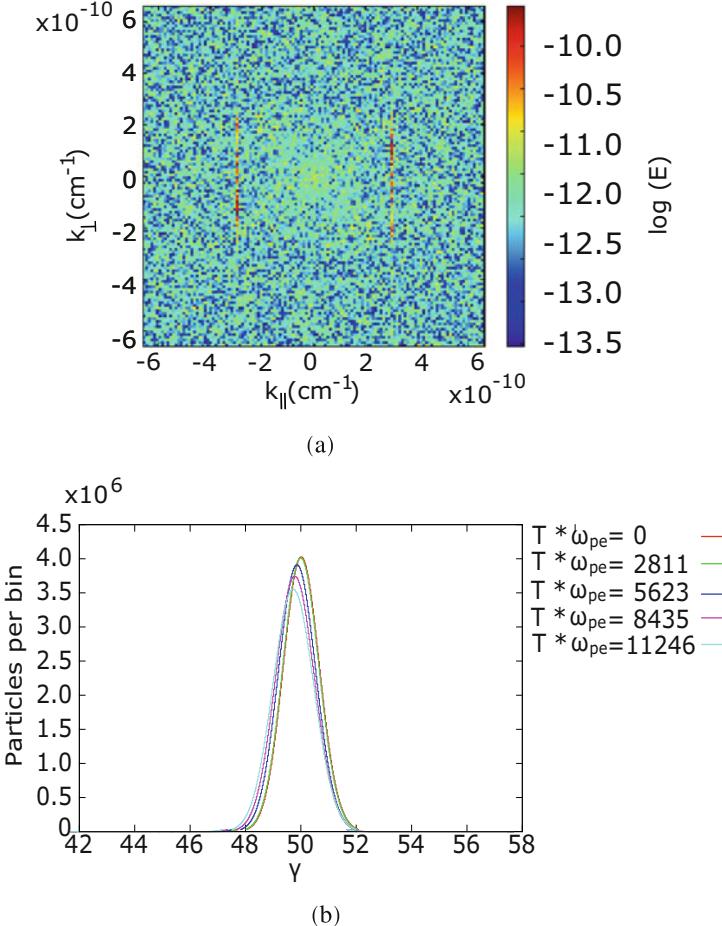


Fig. 6.12 Simulation of propagating pair beams with $\epsilon = 1$, $\alpha = 2.5 \cdot 10^{-5}$, $\gamma = 50$. (a) E-field 2D Fourier at $t = 11246\omega_{pe}^{-1}$. (b) Particle histogram at different times

stream), from which we find the background density ratio $n_{jet}/n_{bg} = 2/3$ and the ratio $n_{p^+}/n_{e^+} = 1$ in the pair-proton stream. In the lab frame (the rest frame of the simulation box), the two streams are counterstreaming along the z-direction with a Lorentz factor $\gamma = 10$ ($\beta = v/c = 0.995$) each, the electron distribution has a thermal velocity of $v_{th,e} = 0.1c$ in every direction in the rest frame of the moving medium, and the thermal velocity of the protons is $v_{th,p} = 0.1c \cdot (m_p/m_e)^{-1}$. The simulation is initialized by drawing random particles from the thermal distribution and boosting them into the frame of the respective streaming population.

Three-dimensional simulations with five different compositions of counterstreaming plasmas using $128 \times 128 \times 512$ cells with a total of 167 million particles

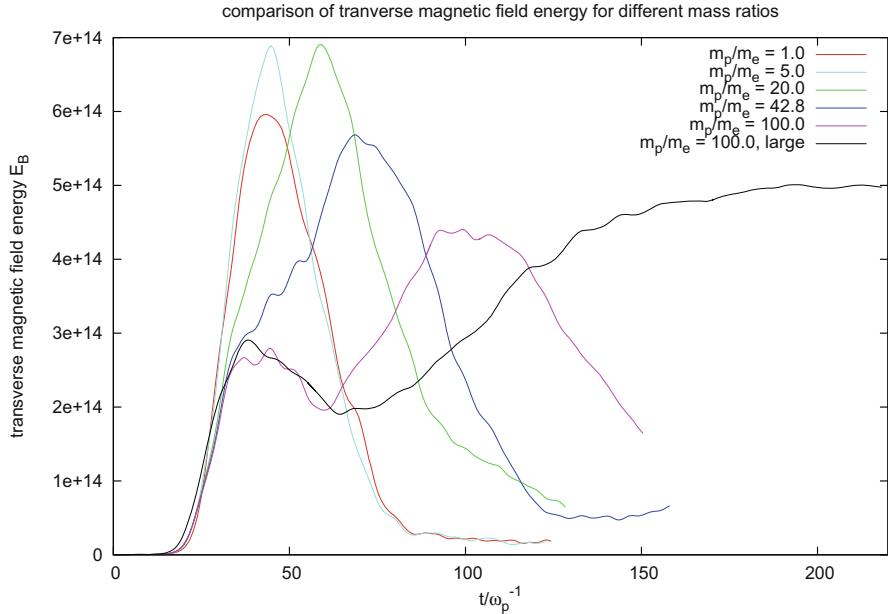


Fig. 6.13 Comparison of the time evolution of the transverse magnetic field energy for different mass ratios. With increasing mass ratio, the instability takes longer to develop. For the two highest mass ratios, one notices that the instability develops in two phases (Figure from [128] “Thomas Burkart, Oliver Elbracht, Urs Ganse, and Felix Spanier: The Influence of the Mass Ratio on the Acceleration of Particles by Filamentation Instabilities,-Astrophysical Journal, Volume 720, Issue 2, pp. 1318-1324.” Published 2010 August 19. Reprinted with permission by ©AAS)

(20 particles per cell) and mass ratios m_p/m_e between 1 and 100 have been performed.

Running and analyzing the simulations are rather similar to the recipe described above. In first place the energy within the simulation should be carefully analyzed. This has been done for the transverse magnetic field energy in Fig. 6.13. The filamentation instability produces a magnetic field in an initially unmagnetized medium. This can be nicely observed here. The statement about linear and nonlinear stages of the instability is also true. In the first stage, there is an almost exponential rise of the magnetic field energy (as expected for a constant growth rate), and then a linear rise follows until the instability comes to a stop. The most interesting feature seen here is that the filaments decay and another instability shows up.

The corresponding filaments are shown in Fig. 6.14. Quite nicely the fine filaments are seen at medium times. These decay at later times as expected. The filaments are best shown in a real space cut orthogonal to the streaming direction.

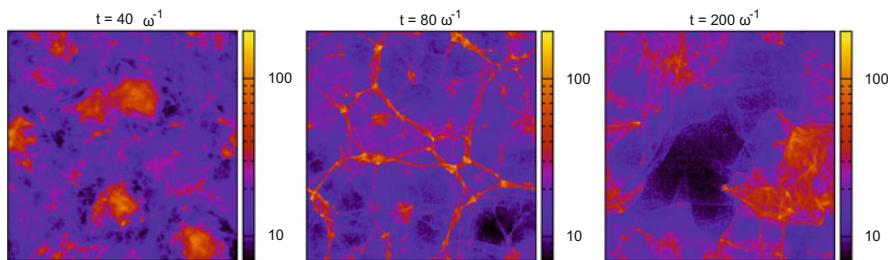


Fig. 6.14 Development and merging of the flux tubes for a simulation with $m_p/m_e = 100.0$ and a resolution of $256 \times 256 \times 512$ cells. The colors pertain to the particle density (in particles per cell). In the three pictures, it is to be seen that first the flux tubes develop, later they nearly vanish (which corresponds to the dip in the magnetic field energy in Fig. 6.13 in the black curve at $80\omega_p^{-1}$), and then the flux tube in the lower right corner grows stronger (lower set of pictures). This can be attributed to the independent instabilities at high mass ratios (Figure from [128] “Thomas Burkart, Oliver Elbracht, Urs Ganse, and Felix Spanier: The Influence of the Mass Ratio on the Acceleration of Particles by Filamentation Instabilities,-Astrophysical Journal, Volume 720, Issue 2, pp. 1318–1324.” Published 2010 August 19. Reprinted with permission by ©AAS)

6.3.4 Collisionless Shocks

There is a long and fruitful history of using particle-in-cell simulations to study the behavior of collisionless shocks. It is such an important topic that the book contains the dedicated section “Astrophysical ShockWaves”. In particular we recommend the book [133] by David Burgess.

6.3.5 Reconnection

Reconnection might be the most studied process in collisionless plasmas, and particle-in-cell codes are often used as no assumptions on underlying resistivity have to be made. This book contains several other chapters where magnetic reconnection is discussed in detail, in particular the Chap. 1.4.1 (reconnection in Kelvin-Helmholtz), Chap. 2.4.4 (reconnection in Hall-MHD), Chap. 3.3 (hybrid simulation, specifically Chap. 3.3.2 “Closure models” and Chap. 3.3.3 with an explicit comparison to PiC), Chap. 5.6.2.1 (reconnection in turbulence), Chap. 7.6.4 (reconnection in flux-transfer-events), Chap. 9.5.1 (reconnection as an application of hybrid codes) and Chap. 11.3 (astrophysical reconnection).

References

1. T. Takizuka, H. Abe, J. Comput. Phys. **25**(3), 205 (1977). [https://doi.org/10.1016/0021-9991\(77\)90099-7](https://doi.org/10.1016/0021-9991(77)90099-7)
2. R.H. Miller, M.R. Combi, Geophys. Res. Lett. **21**(16), 1735 (1994). <https://doi.org/10.1029/94GL01835>

3. C.S. Meierbachtol, A.D. Greenwood, J.P. Verboncoeur, B. Shanker, IEEE Trans. Plasma Sci. **43**(11), 3778 (2015). <https://doi.org/10.1109/TPS.2015.2487522>
4. W. Daughton, J. Scudder, H. Karimabadi, Phys. Plasmas **13**(7), 072101 (2006). <https://doi.org/10.1063/1.2218817>
5. D. Eremin, T. Hemke, R.P. Brinkmann, T. Mussenbrock, J. Phys. D: Appl. Phys. **46**(8), 084017 (2013)
6. J.W. Eastwood, Comput. Phys. Commun. **64**(2), 252 (1991). [https://doi.org/10.1016/0010-4655\(91\)90036-K](https://doi.org/10.1016/0010-4655(91)90036-K)
7. K. Yee, IEEE Trans. Antennas Propag. **14**(3), 302 (1966)
8. J. Räbinä, On a numerical solution of the maxwell equations by discrete exterior calculus. Ph.D. Thesis, University of Jyväskylä (2014)
9. D.Y. Na, B.H.V. Borges, F.L. Teixeira, J. Comput. Phys. **376**, 249 (2019). <https://doi.org/10.1016/j.jcp.2018.09.024>
10. T.D. Arber, K. Bennett, C.S. Brady, A. Lawrence-Douglas, M.G. Ramsay, N.J. Sircombe, P. Gillies, R.G. Evans, H. Schmitz, A.R. Bell, C.P. Ridgers, Plasma Phys. Controlled Fusion **57**(11), 113001 (2015). <https://doi.org/10.1088/0741-3335/57/11/113001>
11. K.J. Bowers, B.J. Albright, L. Yin, B. Bergen, T.J.T. Kwan, Phys. Plasmas **15**(5), 055703 (2008). <https://doi.org/10.1063/1.2840133>
12. A. Seaton, S. Luedtke, Units in vpic (2020). <https://github.com/lanl/vpic/blob/master/doc/vpicUnits.pdf>
13. H. Okuda, J. Comput. Phys. **10**(3), 475 (1972). [https://doi.org/10.1016/0021-9991\(72\)90048-4](https://doi.org/10.1016/0021-9991(72)90048-4)
14. M. Shalaby, A.E. Broderick, P. Chang, C. Pfommer, A. Lamberts, E. Puchwein, Astrophys. J. **841**(1), 52 (2017). <https://doi.org/10.3847/1538-4357/aa6d13>
15. A. Friedman, J. Comput. Phys. **90**(2), 292 (1990). [https://doi.org/10.1016/0021-9991\(90\)90168-Z](https://doi.org/10.1016/0021-9991(90)90168-Z)
16. R.W. Hockney, J.W. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, New York, 1981)
17. C.K. Birdsall, A.B. Langdon, *Plasma Physics via Computer Simulation*, 1st edn. (Taylor and Francis, New York, 2005)
18. B.B. Godfrey, J. Comput. Phys. **15**(4), 504 (1974)
19. J. Fang, Time domain finite difference computation for maxwell's equations. Ph.D. Thesis, University of California, Berkely (1989)
20. M. Hadi, M. Piket-May, IEEE Trans. Antennas Propag. **45**(2), 254 (1997). <https://doi.org/10.1109/8.560344>
21. A.D. Greenwood, K.L. Cartwright, J.W. Luginsland, E.A. Baca, J. Comput. Phys. **201**(2), 665 (2004). <https://doi.org/10.1016/j.jcp.2004.06.021>
22. J. Cole, IEEE Trans. Antennas Propag. **50**(9), 1185 (2002). <https://doi.org/10.1109/TAP.2002.801268>
23. M. Kärkkäinen, E. Gjonaj, T. Lau, T. Weiland, in *Proceedings of ICAP 2006*, vol. 1 (Joint Accelerator Conferences Website, Chamonix, France, 2006), vol. 1, p. 35
24. J.L. Vay, C. Geddes, E. Cormier-Michel, D. Grote, J. Comput. Phys. **230**(15), 5908 (2011). <https://doi.org/10.1016/j.jcp.2011.04.003>
25. I. Haber, R. Lee, H. Klein, J. Boris, in *Proc. Sixth Conf. Num. Sim. Plasmas, Berkeley, CA* (1973), pp. 46–48
26. S. Jalas, I. Dornmair, R. Lehe, H. Vincenti, J.L. Vay, M. Kirchen, A.R. Maier, Phys. Plasmas **24**(3), 033115 (2017). <https://doi.org/10.1063/1.4978569>
27. J.L. Vay, I. Haber, B.B. Godfrey, J. Comput. Phys. **243**, 260 (2013). <https://doi.org/10.1016/j.jcp.2013.03.010>
28. R.W. Hockney, Phys. Fluids **9**(9), 1826 (1966). <https://doi.org/10.1063/1.1761939>
29. S.J. Gitomer, Phys. Fluids **15**(10), 1861 (1972). <https://doi.org/10.1063/1.1693791>
30. C.G. Darwin, Philos. Mag. Series 5 **39** (1920). <https://doi.org/10.1080/14786440508636066>
31. C.K. Birdsall, D. Fuss, J. Comput. Phys. **135**(2), 141 (1968). <https://doi.org/10.1006/jcph.1997.5723>

32. A. Langdon, J. Comput. Phys. **6**(2), 247 (1970). [https://doi.org/10.1016/0021-9991\(70\)90024-0](https://doi.org/10.1016/0021-9991(70)90024-0)
33. B.A. Shadwick, C.B. Schroeder, AIP Conf. Proc. **1086**(1), 321 (2009). <https://doi.org/10.1063/1.3080926>
34. J. Eastwood, R. Hockney, J. Comput. Phys. **16**(4), 342 (1974). [https://doi.org/10.1016/0021-9991\(74\)90044-8](https://doi.org/10.1016/0021-9991(74)90044-8)
35. Y. Matsuda, H. Okuda, Phys. Fluids **18**(12), 1740 (1975). <https://doi.org/10.1063/1.861092>
36. I.V. Sokolov, Comput. Phys. Commun. **184**(2), 320 (2013). <https://doi.org/10.1016/j.cpc.2012.09.015>
37. H. Abe, N. Sakairi, R. Itatani, H. Okuda, J. Comput. Phys. **63**(2), 247 (1986). [https://doi.org/10.1016/0021-9991\(86\)90193-2](https://doi.org/10.1016/0021-9991(86)90193-2)
38. P. Kilian, U. Ganse, F. Spanier, in *Numerical Modeling of Space Plasma Flows (ASTRONUM2012)*, *Astronomical Society of the Pacific Conference Series*, vol. 474, ed. by N.V. Pogorelov, E. Audit, G.P. Zank (2013). Astronomical Society of the Pacific Conference Series, vol. 474, p. 208
39. P.A. Muñoz, P. Kilian, J. Büchner, Phys. Plasmas **21**(11), 112106 (2014). <https://doi.org/10.1063/1.4901033>
40. W. Wu, H. Qin, Phys. Plasmas **25**(10), 102107 (2018). <https://doi.org/10.1063/1.5038039>
41. J.P. Boris, in *Proceedings of the Fourth Conference on the Numerical Simulation of Plasmas, Washington DC*, ed. by J. Boris, R. Shanny (Naval Research Laboratory, Washington DC, 1970), pp. 3–67
42. J. Vay, Phys. Plasmas **15**(5), 056701 (2008). <https://doi.org/10.1063/1.2837054>
43. A.V. Higuera, J.R. Cary, Phys. Plasmas **24**(5), 052104 (2017). <https://doi.org/10.1063/1.4979989>
44. T. Umeda, Comput. Phys. Commun. **228**, 1 (2018). <https://doi.org/10.1016/j.cpc.2018.03.019>
45. B. Ripperda, F. Bacchini, J. Teunissen, C. Xia, O. Porth, L. Sironi, G. Lapenta, R. Keppens, Astrophys. J. Suppl. Series **235**(1), 21 (2018). <https://doi.org/10.3847/1538-4365/aab114>
46. S. Zenitani, T. Umeda, Phys. Plasmas **25**(11), 112110 (2018). <https://doi.org/10.1063/1.5051077>
47. B. Marder, J. Comput. Phys. **68**, 48 (1987). [https://doi.org/10.1016/0021-9991\(87\)90043-X](https://doi.org/10.1016/0021-9991(87)90043-X)
48. J. Villasenor, O. Buneman, Comput. Phys. Commun. **69**(2–3), 306 (1992). [https://doi.org/10.1016/0010-4655\(92\)90169-Y](https://doi.org/10.1016/0010-4655(92)90169-Y)
49. T.Z. Esirkepov, Comput. Phys. Commun. **135**(2), 144 (2001). [https://doi.org/10.1016/S0010-4655\(00\)00228-9](https://doi.org/10.1016/S0010-4655(00)00228-9)
50. T. Umeda, Y. Omura, T. Tominaga, H. Matsumoto, Comput. Phys. Commun. **156**(1), 73 (2003). [https://doi.org/10.1016/S0010-4655\(03\)00437-5](https://doi.org/10.1016/S0010-4655(03)00437-5)
51. J.H. Halton, Commun. ACM **7**(12), 701 (1964). <https://doi.org/10.1145/355588.365104>
52. J.A. Byers, M. Grewal, Phys. Fluids **13**(7), 1819 (1970). <https://doi.org/10.1063/1.1693160>
53. M. Melzani, C. Winisdoerffer, R. Walder, D. Folini, J.M. Favre, S. Krastanov, P. Messmer, Astron. Astrophys. **558**, A133 (2013). <https://doi.org/10.1051/0004-6361/201321557>
54. S. Zenitani, Phys. Plasmas **22**(4), 042116 (2015). <https://doi.org/10.1063/1.4919383>
55. K. Cartwright, J. Verboncoeur, C. Birdsall, J. Comput. Phys. **162**(2), 483 (2000). <https://doi.org/10.1006/jcph.2000.6549>
56. D. Welch, T. Genoni, R. Clark, D. Rose, J. Comput. Phys. **227**(1), 143 (2007). <https://doi.org/10.1016/j.jcp.2007.07.015>
57. D. Faghihi, V. Carey, C. Michoski, R. Hager, S. Janhunen, C.S. Chang, R. Moser (2017). arXiv e-prints arXiv:1702.05198
58. R. Martin, J.L. Cambier, in *28th International Symposium on Rarefied Gas Dynamics* (AIP Publishing, Zaragoza, Spain, 2012), AIP Conference Proceedings. <https://doi.org/10.1063/1.4769634>
59. M. Vranic, T. Grismayer, J. Martins, R. Fonseca, L. Silva, Comput. Phys. Commun. **191**, 65 (2015). <https://doi.org/10.1016/j.cpc.2015.01.020>
60. D. Bruhwiler, R. Giaccone, J. Cary, J. Verboncoeur, P. Mardahl, E. Esarey, W. Leemans, B. Shadwick, Phys. Rev. Special Topics Accl. Beams **4**(10), 1 (2001)

61. P. Yu, X. Xu, A. Davidson, A. Tableman, T. Dalichaouch, F. Li, M.D. Meyers, W. An, F.S. Tsung, V.K. Decyk, F. Fiuzza, J. Vieira, R.A. Fonseca, W. Lu, L.O. Silva, W.B. Mori, *J. Comput. Phys.* **316**, 747 (2016). <https://doi.org/10.1016/j.jcp.2016.04.014>
62. Z.P. Liao, H.L. Wong, B.P. Yang, Y.F. Yuan, *Scientia Sinica (Series A)* **17**, 1063 (1984)
63. G. Mur, *IEEE Trans. Electromagn. Compat.* **EMC-23**(4), 377 (1981)
64. A. Tafove, S.C. Hagness, *Computational Electrodynamics: The Finite-Differences Time-Domain Method*, 2nd edn. (Artech House, Boston, London, 2000)
65. J.P. Berenger, *J. Comput. Phys.* **114**(2), 185 (1994)
66. J.P. Berenger, *Synth. Lect. Comput. Electromagn.* **2**(1), 1 (2007). <https://doi.org/10.2200/S00030ED1V01Y200605CEM008>
67. A. Klimas, M. Hesse, S. Zenitani, *Phys. Plasmas* **15**(8), 082102 (2008). <https://doi.org/10.1063/1.2965826>
68. H. Ohtani, R. Horiuchi, *Plasma Fusion Res.* **4**, 024 (2009). <https://doi.org/10.1585/pfr.4.024>
69. I. Thiele, S. Skupin, R. Nuter, *J. Comput. Phys.* **321**, 1110 (2016). <https://doi.org/10.1016/j.jcp.2016.06.004>
70. K. Ardaneh, R. Giust, B. Morel, F. Courvoisier, *Opt. Exp.* **28**(3), 2895 (2020). <https://doi.org/10.1364/OE.385413>
71. F. Pérez, M. Grech (2018). arXiv e-prints arXiv:1809.04435
72. P. Kilian, F. Spanier, *J. Comput. Phys.* **353**, 258 (2018). <https://doi.org/10.1016/j.jcp.2017.10.012>
73. H.C. Edwards, C.R. Trott, D. Sunderland, *J. Parallel Distrib. Comput.* **74**(12), 3202 (2014). <https://doi.org/10.1016/j.jpdc.2014.07.003>. Domain-Specific Languages and High-Level Frameworks for High-Performance Computing
74. D.A. Beckingsale, J. Burmark, R. Hornung, H. Jones, W. Killian, A.J. Kunen, O. Pearce, P. Robinson, B.S. Ryujin, T.R. Scogland, in *2019 IEEE/ACM International Workshop on Performance, Portability and Productivity in HPC (p3hpc)* (IEEE, 2019), pp. 71–81
75. L. Dagum, R. Menon, *IEEE Comput. Sci. Eng.* **5**(1), 46 (1998). <https://doi.org/10.1109/99.660313>
76. N.M. Larsgård, S. Anne, C. Elster, Parallelizing particle-in-cell codes with openmp and mpi (2007). <http://citeserx.ist.psu.edu/viewdoc/summary?doi=10.1.1.216.342>
77. D.W. Walker, J.J. Dongarra, *Supercomputer* **12**, 56 (1996)
78. M. Bauer, S. Treichler, E. Slaughter, A. Aiken, in *SC'12: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis* (IEEE, 2012), pp. 1–11
79. R.E.W. Pfund, R. Lichters, J. Meyer-ter Vehn, *AIP Conf. Proc.* **426**(1), 141 (1998). <https://doi.org/10.1063/1.55199>
80. S.J. Plimpton, D.B. Seidel, M.F. Pasik, R.S. Coats, Load-balancing techniques for a parallel electromagnetic particle-in-cell code. Tech. rep., Sandia National Labs (2000). <https://doi.org/10.2172/751032>
81. J. Qiang, R.D. Ryne, S. Habib, V. Decyk, *J. Comput. Phys.* **163**(2), 434 (2000). <https://doi.org/10.1006/jcph.2000.6570>
82. P. Messmer, in *Applied Parallel Computing. New Paradigms for HPC in Industry and Academia*, ed. by T. Sørevik, F. Manne, A.H. Gebremedhin, R. Moe (Springer Berlin Heidelberg, Berlin, Heidelberg, 2001), pp. 350–355
83. J. Qiang, X. Li, *Comput. Phys. Commun.* **181**(12), 2024 (2010). <https://doi.org/10.1016/j.cpc.2010.08.021>
84. A. Beck, J. Frederiksen, J. Dérouillat, *Nucl. Instrum. Methods Phys. Res. Sect. A: Accel. Spectrom. Detectors Associated Equip.* **829**, 418 (2016). <https://doi.org/10.1016/j.nima.2016.03.112>. 2nd European Advanced Accelerator Concepts Workshop—EAAC 2015
85. K. Germaschewski, W. Fox, S. Abbott, N. Ahmadi, K. Maynard, L. Wang, H. Ruhl, A. Bhattacharjee, *J. Comput. Phys.* **318**, 305 (2016). <https://doi.org/10.1016/j.jcp.2016.05.013>
86. Y. Barsamian, S.A. Hirstoaga, Éric Violard, *J. Comput. Sci.* **27**, 345 (2018). <https://doi.org/10.1016/j.jocs.2018.06.004>

87. K.G. Miller, R.P. Lee, A. Tableman, A. Helm, R.A. Fonseca, V.K. Decyk, W.B. Mori, Comput. Phys. Commun. **259**, 107633 (2021). <https://doi.org/10.1016/j.cpc.2020.107633>
88. J. Kaiser, R. Schafer, IEEE Trans. Acoust. Speech Signal Proc. **28**(1), 105 (1980). <https://doi.org/10.1109/TASSP.1980.1163349>
89. R. Pausch, A. Debus, A. Huebl, U. Schramm, K. Steiniger, R. Widera, M. Bussmann, Nucl. Instrum. Methods Phys. Res. Sect. A: Accel. Spectrom. Detectors Associated Equip. **909**, 419 (2018). <https://doi.org/10.1016/j.nima.2018.02.020>. 3rd European Advanced Accelerator Concepts workshop (EAAC2017)
90. E.N. Nerush, D.A. Serebryakov, I.Y. Kostyukov, Astrophys. J. **851**(2), 129 (2017). <https://doi.org/10.3847/1538-4357/aa9d1a>
91. C. Kalapotharakos, G. Brambilla, A. Timokhin, A.K. Harding, D. Kazanas, Astrophys. J. **857**(1), 44 (2018). <https://doi.org/10.3847/1538-4357/aab550>
92. G. Jacobs, J. Hesthaven, J. Comput. Phys. **214**(1), 96 (2006). <https://doi.org/10.1016/j.jcp.2005.09.008>
93. P.A. Delamere, R.J. Wilson, A. Masters, J. Geophys. Res. Space Phys. **116**(A10) (2011). <https://doi.org/10.1029/2011JA016724>
94. T. Amano, K. Higashimori, K. Shirakawa, J. Comput. Phys. **275**, 197 (2014). <https://doi.org/10.1016/j.jcp.2014.06.048>
95. P. Muñoz, N. Jain, P. Kilian, J. Büchner, Comput. Phys. Commun. **224**, 245 (2018). <https://doi.org/10.1016/j.cpc.2017.10.012>
96. P. Kilian, P.A. Muñoz, C. Schreiner, F. Spanier, J. Plasma Phys. **83**(1), 707830101 (2017). <https://doi.org/10.1017/S0022377817000149>
97. E. Åström, *On Waves in an Ionized Gas*. Arkiv för fysik (Almqvist & Wiksell, 1950)
98. I.B. Bernstein, Phys. Rev. **109**, 10 (1958). <https://doi.org/10.1103/PhysRev.109.10>
99. M. André, J. Plasma Phys. **33**, 1 (1985). <https://doi.org/10.1017/S0022377800002270>
100. D. Swanson, *Plasma Waves*, 2nd edn. Series in Plasma Physics (Taylor & Francis, 2003)
101. J.P. Dougherty, D.T. Farley, R. Soc. Lond. Proc. Ser. A **259**, 79 (1960). <https://doi.org/10.1098/rspa.1960.0212>
102. A. Akhiezer, I. Akhiezer, R. Polovin, A. Sitenko, K. Stepanov, *Plasma Electrodynamics*. International Series in Natural Philosophy, vol. 68 (Pergamon, 1975). <https://doi.org/10.1016/B978-0-08-017783-0.50007-5>
103. P.H. Yoon, R.A. López, Phys. Plasmas **24**(2), 022117 (2017). <https://doi.org/10.1063/1.4976321>
104. K. Roennmark, Waves in homogeneous, anisotropic multicomponent plasmas (WHAMP). Tech. rep., Kiruna Geophysical Institute (1982)
105. H. Xie, Y. Xiao, Plasma Sc. Technol. **18**(2), 97 (2016). <https://doi.org/10.1088/1009-0630/18/2/01>
106. J.M. Dawson, Phys. Fluids **4**(7), 869 (1961). <https://doi.org/10.1063/1.1706419>
107. G. Manfredi, Phys. Rev. Lett. **79**, 2815 (1997). <https://doi.org/10.1103/PhysRevLett.79.2815>
108. A.M. Dimits, G. Bateman, M.A. Beer, B.I. Cohen, W. Dorland, G.W. Hammett, C. Kim, J.E. Kinsey, M. Kotschenreuther, A.H. Kritz, L.L. Lao, J. Mandrekas, W.M. Nevins, S.E. Parker, A.J. Redd, D.E. Shumaker, R. Sydora, J. Weiland, Phys. Plasmas **7**(3), 969 (2000). <https://doi.org/10.1063/1.873896>
109. P.A. Muñoz, D. Told, P. Kilian, J. Büchner, F. Jenko, Phys. Plasmas **22**(8), 082110 (2015). <https://doi.org/10.1063/1.4928381>
110. G.L. Falchetto, B.D. Scott, P. Angelino, A. Bottino, T. Dannert, V. Grandgirard, S. Janhunen, F. Jenko, S. Jolliet, A. Kndl, B.F. McMillan, V. Naulin, A.H. Nielsen, M. Ottaviani, A.G. Peeters, M.J. Pueschel, D. Reiser, T.T. Ribeiro, M. Rmanelli, Plasma Phys. Controlled Fusion **50**(12), 124015 (2008)
111. F. Riva, C.F. Beadle, P. Ricci, Phys. Plasmas **24**(5), 055703 (2017). <https://doi.org/10.1063/1.4977917>
112. D. Biskamp, E. Schwarz, A. Zeiler, A. Celani, J.F. Drake, Phys. Plasmas **6**, 751 (1999). <https://doi.org/10.1063/1.873312>

-
113. E. Papini, L. Franci, S. Landi, A. Verdini, L. Matteini, P. Hellinger, *Astrophys. J.* **870**, 52 (2019). <https://doi.org/10.3847/1538-4357/aaf003>
 114. D. Groselj, C.H.K. Chen, A. Mallet, R. Samtaney, K. Schneider, F. Jenko (2018). arXiv e-prints
 115. C. Schreiner, F. Spanier, *Comput. Phys. Commun.* **185**, 1981 (2014). <https://doi.org/10.1016/j.cpc.2014.03.028>
 116. S.S. Cerri, L. Franci, F. Califano, S. Landi, P. Hellinger, *J. Plasma Phys.* **83**(2), 705830202 (2017). <https://doi.org/10.1017/S0022377817000265>
 117. D. Grošelj, S.S. Cerri, A. Bañón Navarro, C. Willmott, D. Told, N.F. Loureiro, F. Califano, F. Jenko, *Astrophys. J.* **847**, 28 (2017). <https://doi.org/10.3847/1538-4357/aa894d>
 118. V. Zhdankin, G.R. Werner, D.A. Uzdensky, M.C. Begelman, *Phys. Rev. Lett.* **118**(5), 055103 (2017). <https://doi.org/10.1103/PhysRevLett.118.055103>
 119. L. Comisso, L. Sironi, *Phys. Rev. Lett.* **121**(25), 255101 (2018). <https://doi.org/10.1103/PhysRevLett.121.255101>
 120. R. Schlickeiser, *Cosmic Ray Astrophysics* (Springer, Berlin, 2002)
 121. M. Wisniewski, F. Spanier, R. Kissmann, *Astrophys. J.* **750**, 150 (2012). <https://doi.org/10.1088/0004-637X/750/2/150>
 122. S. Lange, F. Spanier, M. Battarbee, R. Vainio, T. Laitinen, *Astron. Astrophys.* **553**, A129 (2013). <https://doi.org/10.1051/0004-6361/201220804>
 123. A. Ivascenko, S. Lange, F. Spanier, R. Vainio, *Astrophys. J.* **833**, 223 (2016). <https://doi.org/10.3847/1538-4357/833/2/223>
 124. R.L. Morse, C.W. Nielson, *Phys. Fluids* **14**, 830 (1971). <https://doi.org/10.1063/1.1693518>
 125. A. Bret, *Astrophys. J.* **699**, 990 (2009). <https://doi.org/10.1088/0004-637X/699/2/990>
 126. A. Bret, *Laser Particle Beams* **28**, 491 (2010). <https://doi.org/10.1017/S0263034610000480>
 127. M. Shalaby, A.E. Broderick, P. Chang, C. Pfrommer, A. Lamberts, E. Puchwein, *Astrophys. J.* **848**, 81 (2017). <https://doi.org/10.3847/1538-4357/aa8b17>
 128. T. Burkart, O. Elbracht, U. Ganse, F. Spanier, *Astrophys. J.* **720**, 1318 (2010). <https://doi.org/10.1088/0004-637X/720/2/1318>
 129. Q. Moreno, M.E. Dieckmann, X. Ribeyre, S. Jequier, V.T. Tikhonchuk, E. d'Humières, *Phys. Plasmas* **25**, 062125 (2018). <https://doi.org/10.1063/1.5027913>
 130. A. Neronov, I. Vovk, *Science* **328**, 73 (2010). <https://doi.org/10.1126/science.1184192>
 131. L. Sironi, D. Giannios, *Astrophys. J.* **787**, 49 (2014). <https://doi.org/10.1088/0004-637X/787/1/49>
 132. A. Kempf, P. Kilian, F. Spanier, *Astron. Astrophys.* **585**, A132 (2016). <https://doi.org/10.1051/0004-6361/201527521>
 133. D. Burgess, M. Scholer, *Collisionless Shocks in Space Plasmas: Structure and Accelerated Particles*. Cambridge Atmospheric and Space Science Series (Cambridge University Press, Cambridge, 2015). <https://doi.org/10.1017/CBO9781139044097>

Part II

Introduction to Advanced Simulation Approaches

Not all plasma-astrophysical processes can be treated by the methods described in the first part of this book. Advanced simulation methods have been developed, which offer ways to treat open problems. Some of the most important of these methods are described in the second part of this book. One major problem of simulating space and astrophysical plasma phenomena is the large range of scales over which they develop. Chapter 7 shows how adaptive methods allow very global MHD simulations by adjusting the numerical resolution of the relevant (fluid-) processes. Other problems arise in the description of such important phenomena, such as shock waves, magnetic reconnection, and turbulence that are scale-overarching from kinetic to fluid processes. As a result, small-scale kinetic processes that are not tractable in the MHD approximation might be important to understand the large-scale system evolution of a fluid. Unfortunately, the spatial and temporal extension of kinetic plasma simulations is limited to rather small spatial domains and short durations. To consider the consequences of microscopic processes for the large-scale plasma dynamics, multi-scale kinetic simulations have been developed, which are presented in Chap. 8. To kinetically describe at least the ions but considering electrons as a fluid, Chaps. 3 and 5 introduced hybrid-kinetic models and algorithms. They traditionally, however, completely neglect the electron mass to save computer time. Modern parallel-calculating, high-performance supercomputers nowadays allow the full consideration of the electron inertia by hybrid codes. Chapter 9 presents appropriate algorithms for that sake and demonstrates the necessity of considering the electron mass to properly understand reconnection, turbulence, and other processes. Chapter 10 shows how even ions heavier than protons can be treated as particles in hybrid-kinetic simulations, if one uses generalized quasi-neutral schemes. Astrophysical applications of the hot plasma simulation methods developed for space physics are discussed in Chap. 11, where it is shown that astrophysical shock waves, magnetic reconnection, the acceleration of cosmic rays as well as the dynamics of accretion discs and their magneto-rotational instability can be modeled successfully.



Adaptive Global Magnetohydrodynamic Simulations

7

Tamas I. Gombosi, Yuxi Chen, Zhenguang Huang,
Ward B. Manchester, Igor Sokolov, Gabor Toth,
and Bart van der Holst

Abstract

This chapter briefly summarizes global MHD codes that are presently used in geospace modeling with particular attention to their availability for community use through the Community Coordinated Modeling Center (CCMC). It describes the underlining methodologies of modern codes using the University of Michigan's Space Weather Modeling Framework (SWMF) and the Block-Adaptive Tree Solar-Wind Roe-Type Upwind Scheme (BATS-R-US) multi-physics code. We also describe the new MHD with embedded PIC (MHD-EPIC) code, the latest development in adaptive physics approach.

7.1 Introduction

Numerical simulation and modeling have become increasingly essential to basic and applied space physics research for two primary reasons. First, the heliosphere and magnetosphere are vast regions of space with relatively few *in situ* measurements. To understand the global behavior of this complex system, numerical simulations provide the ability to “stitch together” observations from different regions to provide insight to the interpretation of data. The second reason for the increased reliance upon simulations is that the models themselves have evolved to a point where their physical content and numerical robustness, flexibility, and improving ease of

T. I. Gombosi (✉) · Y. Chen · Z. Huang · W. B. Manchester · I. Sokolov · G. Toth ·

B. Van der Holst

Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

e-mail: amas@umich.edu; yuxichen@umich.edu; zghuang@umich.edu; igorsok@umich.edu;
gtoth@umich.edu; bartvand@umich.edu

use inspire researchers to apply them to intriguing scenarios with a new measure of confidence. However, many shortcomings and questions remain for even the most advanced models in terms of inclusion of important physical mechanisms, the spatial and temporal domains that can be addressed, and thorny technical numerical issues to be solved. Nonetheless, it can be safely stated that modeling has over the last several years crossed a threshold whereby they have made the transition from the arcane preserves of specialists to practical tools with widespread application.

Global computational models based on first principles mathematical descriptions of the physics represent a very important component of efforts to understand plasma phenomena associated with the solar system including the large-scale solar corona, the solar wind, its interaction with planetary magnetospheres, comets, the interstellar medium, and the initiation, structure, and evolution of solar eruptive events. In our view, presently, and in the foreseeable future, numerical models based on the equations of magnetohydrodynamics (MHD) are the only self-consistent mathematical descriptions that can span the enormous distances associated with large-scale phenomena in space. Although providing only a relatively low-order approximation to the actual behavior of plasmas, MHD models have been used successfully to simulate many important space plasma processes and provide a powerful means for significantly advancing the understanding of such processes.

7.2 Brief History of Global MHD Simulations of Space Plasmas

Global MHD models for space science applications were first developed in the early 1980s [1–4]. In the late 1990s and early 2000s, the performance, the robustness, and the level of sophistication of the global models dramatically improved to the point that simulations became the “third branch” of investigation methods complementing observations and theory. Most of the presently operational models are based on single-fluid ideal MHD with numerical dissipation (resistivity, diffusion, viscosity, etc.) “mimicking” physical processes missing from the governing equations and thus limiting the physical realism of these codes. More advanced “research codes” incorporate physically more accurate descriptions of these processes by using resistive MHD, multispecies and multifluid descriptions (multispecies description uses separate continuity equations for the ion components but assumes that they all have the same bulk velocity and temperature, while multifluid approximation uses separate continuity, momentum, and energy equations for all ion species), the Hall term, anisotropic plasma pressure, and other improvements.

Most global space weather simulation codes were primarily designed to solve the governing equations describing a single physics domain (even though some models were later applied to other domains as well). These include the Lyon-Fedder-Mobarry (LFM) [5, 6], the OpenGGCM [7, 8], the Watanabe-Sato [9, 10], and the GUMICS [11] codes and the Integrated Space Weather Prediction Model (ISM) [12, 13] of the Earth’s magnetosphere, the Magnetohydrodynamics Around a Sphere (MAS) code describing the solar corona [14, 15], Usmanov’s heliosphere code [16, 17], Hayashi’s corona model [18], the SIP-CESE model [19], Odstrcil’s

ENLIL [20, 21] model of the inner heliosphere, and Pogorelov's outer heliosphere model [22]. More general-use models include the widely used ZEUS code [23], Ogino's planetary magnetosphere code [24], Tanaka's 3D global MHD model [25], Winglee's multifluid Hall MHD code [26, 27], Toth's general MHD Versatile Advection Code (VAC) [28], and the University of Michigan's BATS-R-US [29, 30] model. About a decade ago, space weather models started to couple the global magnetohydrodynamic description of large-scale phenomena with regional models describing critical smaller-scale regions or processes. This technology was used for space weather simulations by the University of Michigan's Space Weather Modeling Framework (SWMF) [30, 31], OpenGGCM [32], and the Center for Integrated Space Weather Modeling (CISM) [33, 34].

One of the most recent developments in space weather modeling is the embedding of kinetic simulation regions in global MHD simulations. Two groups initiated this approach: a group located at UCLA used MHD solutions to provide initial and static boundary conditions for a particle-in-cell (PIC) simulation [35], while the University of Michigan group two-way coupled a global Hall MHD code to embedded PIC simulation boxes [36].

This chapter will briefly summarize the equations used by global space weather models and point out the various approximations used by the various codes. We use our BATS-R-US multi-physics extended MHD (X-MHD) model and SWMF to illustrate the various techniques.

7.2.1 Models of the Solar Corona

The first numerical solution of the solar wind equations was carried out by Scarf and Noble [37, 38], who solved the steady-state spherically symmetric Navier-Stokes equations including heat conduction and viscosity effects. Instead of considering a polytropic corona, they solved the energy equation, thus making the problem more challenging. The set of differential equations had a singularity when the local solar wind speed reached the sound speed. In order to avoid numerical problems, Noble and Scarf [37] integrated the equations from the Earth inward. This was made possible by the fact that at that time the solar wind conditions were observed mainly near Earth. A similar solution with heat conduction but without viscosity was also obtained by Whang and Chang [39].

The first two-fluid model of the solar corona (separate electron and proton equations) was published by Sturrock and Hartle [40]. They realized that as the solar wind leaves the vicinity of the Sun, the collisional coupling between electrons and ions becomes weak, and the electron and ion temperatures can deviate from each other. They developed a two-temperature model where the plasma remains quasi-neutral and current-free, but the two temperatures can be different. Sturrock and Hartle [40] still considered a spherically symmetric steady-state problem and neglected magnetic field effects, but their model represented a step forward. By combining the continuity, momentum, and energy equations, they derived separate "heat equations" for electrons and ions [40].

The importance of magnetic fields. By the late 1960s, it has become clear that the magnetic field plays a major role in determining the density as well as the velocity and temperature structure of the corona. The first models used the observed line-of-sight component of the photospheric magnetic field to determine the magnetic field of the solar corona in the current-free (or potential-field) approximation [41–46].

The source surface magnetic field model is based on the fundamental assumption that the magnetic field above the photosphere is current-free and therefore the coronal magnetic field can be represented by a magnetic scalar potential. This potential can be written as an infinite series of spherical harmonics (cf. [47]). It was recognized by Schatten et al. [43] that the coronal magnetic field follows the current-free potential solution between the photosphere and a “source surface” (located at $r = R_s$) where the potential vanishes and the magnetic field becomes radial. This requires a network of thin current sheets for $r \geq R_s$ (cf. [46]). For the optimal radius of the source surface, Schatten et al. [43] found $R_s = 1.6 R_\odot$, while Altschuler and Newkirk [44] recommended $R_s = 2.5 R_\odot$. Today, most potential field source surface (PFSS) models use the $R_s = 2.5 R_\odot$ value. Outside the source surface, it is assumed that the radial flow of the solar wind carries the magnetic field outward into the heliosphere. This region is not described by the PFSS model. The inner boundary condition at the photosphere is obtained from the observed line-of-sight (LOS) magnetic field components using a least square fit (cf. [48]). However, the LOS observation is in a $\sin(\text{latitude})$ grid (typically applied in MDI, HMI, and GONG magnetograms), and it is challenging to obtain the high-order harmonic coefficients (see [49]). Using a finite difference solver avoids those issues. The outer boundary condition at the source surface is that the field is normal to the source surface, consistent with the assumption that it is then carried outward by the solar wind.

Expanding magnetized corona models. Nearly simultaneously with the development of the source surface models, the effect of the radial coronal expansion on the magnetic field was also explored by Pneuman [50–53]. In a series of papers [50–52], he analytically investigated how centrifugal, pressure gradient, and magnetic forces impact the flow of an infinitely conducting fluid where the magnetic field lines and plasma flow lines must coincide in the corotating frame.

An early example of a numerical model of the corona with solar wind is by Pneuman and Kopp [53]. As in the Parker solution [54], this model involves a high temperature and high density plasma whose pressure cannot be held in equilibrium by solar gravity or the pressure of the interstellar medium. Consequently, the coronal plasma expands rapidly outward achieving supersonic speeds within a few solar radii, and in doing so forms the solar wind.

Figure 7.1 shows a comparison of the field lines of the numerical solution and a potential field model with the same normal component at the reference level [53]. In spite of its obvious limitations, the Pneuman and Kopp [53] simulation established the usefulness of MHD simulations of the solar corona. It was the first successful attempt to apply the conservation laws of magnetohydrodynamics to explain large-scale features of the solar corona.

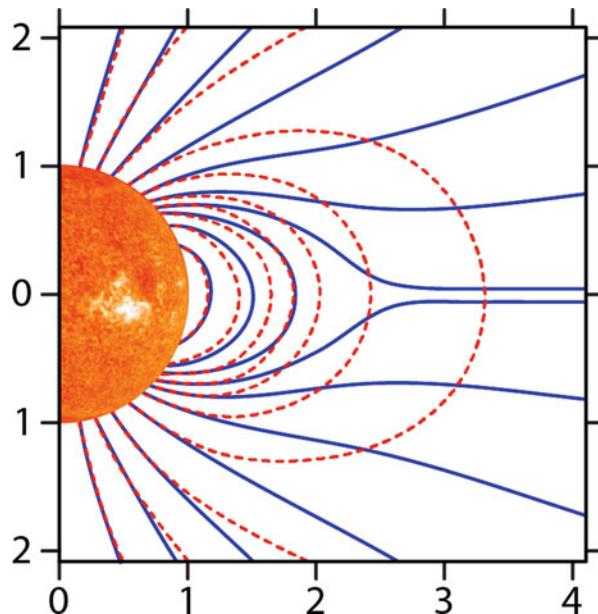


Fig. 7.1 A comparison of the magnetic field (solid blue curves) with a potential field (dashed red curves) having the same normal component at the reference level. To make the comparison meaningful, the field lines for the two cases are chosen so as to be coincident at the surface. As expected, the field is everywhere stretched outward by the gas with this distortion becoming large near the neutral point. In the closed region well below the cusp, the difference between the two configurations is small, mainly because the pressure at the reference level is taken to be independent of latitude (reproduced from [53] with the permission of Springer publishers)

7.2.2 Heliosphere Models

The first generation of magnetohydrodynamic models of the interplanetary medium were developed in the second half of the 1970s and were used for about two decades [55–71]. These models were designed to describe only large-scale bulk-average features of the plasma observed through the solar cycle. At solar minimum, these coronal structures are as follows:

1. Open magnetic field lines forming coronal holes.
2. Closed magnetic field lines forming a streamer belt at low latitudes.
3. The bimodal nature of the solar wind is reproduced with fast wind originating from coronal holes over the poles and slow wind at low latitudes.

A thin current sheet forms at the tip of the streamer belt and separates opposite directed magnetic flux originating from the two poles. At solar minimum, the fast wind lies above 30° heliographic latitude and has an average velocity of 750 km s^{-1} at distances greater than 15 solar radii, at which distance the wind velocity is close

to its terminal velocity. The slow wind, by contrast, is confined close to the global heliospheric current sheet, which lies near the equator at solar minimum. This component of the wind is highly variable, with speeds that lie between 300 and 450 km s⁻¹. The slow solar wind has been suggested by Wang and Sheeley [72] to originate from highly expanding plasma traveling down magnetic flux tubes that originate near coronal hole boundaries. It has also been suggested that the opening of closed flux tubes by interchange reconnection with open flux may release plasma to form the slow solar wind [73–75]. More recent theories have related the slow solar wind to complex magnetic topology flux tubes near the heliospheric current sheet, which are characterized by the squashing factor [76, 77]. At solar maximum, the current sheet is highly inclined with smaller coronal holes forming at all latitudes, while the fast wind is largely absent.

Steinolfson et al. [55] numerically solved the MHD equations for a spherically symmetric (2D) solar corona, neglecting the polar components of velocity and magnetic field. In a follow-up paper, Steinolfson et al. considered a situation when the flow and magnetic field are in the meridional plane [56]. A few years later, Steinolfson et al. [60] revisited the steady global solar corona that was investigated a decade earlier by Pneuman and Kopp [53].

Marching codes. Pizzo took a very different approach to simulating the interplanetary medium [57–59, 63, 65, 67–70]. As a first step, he developed a 3D hydrodynamic model of steady corotating streams in the solar wind [57], assuming a supersonic, inviscid, and polytropic flow beyond approximately 35 R_s . This approach takes advantage of the fact that in the inertial frame, the temporal and azimuthal gradients are related by $\partial_t = -\Omega_\odot \partial_\phi$, where ϕ is the azimuth angle and Ω_\odot is the equatorial angular velocity of the Sun. Pizzo [57, 58] solved the governing equations that describe the dynamical evolution of 3D corotating solar wind structures. The model is limited to those structures that are steady or nearly steady in the frame rotating with the Sun. It utilizes the single-fluid, polytropic, nonlinear, 3D hydrodynamic equations to approximate the dynamics that occur in interplanetary space, where the flow is supersonic and the governing equations are hyperbolic. Later Pizzo [59] extended his marching method to include the interplanetary magnetic field.

Inner heliosphere codes. Pizzo's early work was followed with the development of the ENLIL heliospheric model by Odstrčil [78]. ENLIL has been adapted to accept inner boundary solar wind conditions from a variety of sources, including the Wang-Sheeley-Arge (WSA) model [72, 79] and coronal MHD models [80, 81]. In the case of [82], the inner boundary conditions (outside the critical point) are derived from interplanetary scintillation (IPS) observations [83]. More recently, other groups also developed 3D inner heliosphere models with super-Alfvénic inner boundary conditions using the WSA approach. These include the LFM-Helio model [84, 85] and its successor the GAMERA code [86], the SUSANOO-SW code [87, 88], the MS-FLUKSS suite [89], and EUHFORIA developed by the Leuwen group [90].

Alfvén wave models. Modern solar corona and solar wind codes recognize that Alfvén waves play a major role in the heating and acceleration of the solar wind.

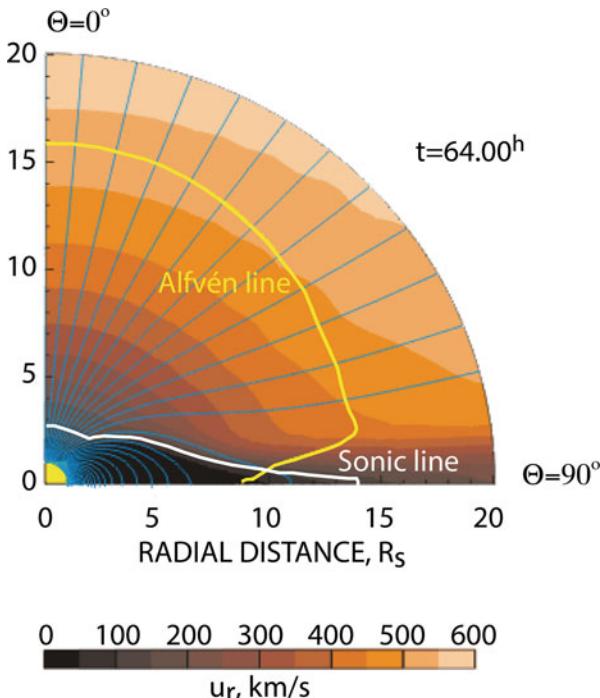


Fig. 7.2 Magnetic field configuration near the Sun superimposed on a map of radial flow velocities after 64 h of relaxation. The Alfvén line (where the radial flow velocity is equal to the Alfvén velocity computed for the total magnetic field) is shown by the yellow line, and the sonic line (where the radial velocity is equal to the sound velocity) is shown by the white line (Reproduced from [17] with the permission of AGU)

This approach was pioneered by Usmanov et al. [17]. Alfvén wave models assume that the generation of reflected counter-propagating waves are the underlying cause of the turbulence energy cascade (e.g., [91]), which transports the energy of turbulence from the large-scale motions across the inertial range of the turbulence spatial scale to short-wavelength perturbations. The latter can be efficiently damped due to wave-particle interactions. In this way, the turbulence energy is converted to random (thermal) energy. Figure 7.2 shows the results of the first axisymmetric (2D) simulation of the solar corona and solar wind using self-consistent Alfvén turbulence [17]. After 64 h of relaxation time, a closed-field region develops near the equator; the flow velocity is high in the open polar field region but decreases toward the equator above the region of the closed magnetic field.

The *ad hoc* elements (like heating functions) of Alfvén wave models can be eliminated by assuming that the coronal plasma is heated by the dissipation of Alfvén wave turbulence [92]. The dissipation itself is caused by the nonlinear interaction between oppositely propagating waves (e.g., [93]). Within coronal holes, there are no closed magnetic field lines; hence, there are no oppositely

propagating waves. Instead, a weak reflection of the outward propagating waves locally generates sunward propagating waves as quantified by van der Holst et al. [94] in their AWSOM code. The small power in these locally generated (and almost immediately dissipated) inward propagating waves leads to a reduced turbulence dissipation rate in coronal holes, naturally resulting in the bimodal solar wind structure. Another consequence is that coronal holes look like cold black spots in the EUV and X-ray images, while closed-field regions are hot and bright. Active regions, where the wave reflection is particularly strong, are the brightest in such models (see [92, 94, 95]). Several recent efforts aim to develop models that include Alfvén waves as a primary driving agent for both heating and accelerating of the solar wind. Examples are [47, 92, 94, 96–102].

7.2.3 Geospace Models

Large-scale MHD simulations of the magnetosphere started at UCLA in the late 1970s under the leadership of John Dawson [1, 2, 103, 104]. The first 2D model [103] was followed by several independent 3D MHD methods employing a Lagrangian approach [1], a Rusanov solver [2], and a Lax-Wendroff technique [104]. These simulations, however, were diffusive and under-resolved to the point that one could hardly recognize the bow shock and the magnetopause (Fig. 7.3).

Early magnetosphere models. The 1963 Treaty Banning Nuclear Weapon Tests in the Atmosphere, in Outer Space and Under Water prohibited all test

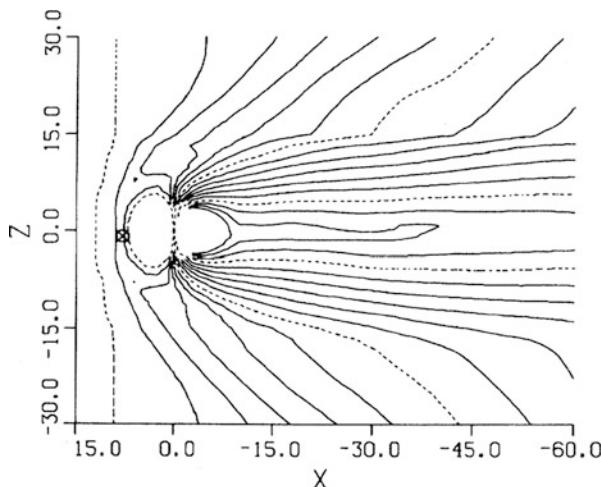


Fig. 7.3 Result of the first “high fidelity” magnetosphere simulation. Plotted are magnetic field lines after a southward interplanetary magnetic field has been incident for 1 h. Contours are equally spaced in vector potential. “x” marks the position of reconnection points (reproduced from [108] with the permission of AIP Publishing)

detonations of nuclear weapons except for those conducted underground. As a result, under the leadership of the Defense Threat Reduction Agency, the United States turned to numerical simulations to study the impacts of high-altitude nuclear explosions. Such simulations need to include realistic models of the ionosphere and magnetosphere as well as the complex physics and chemistry involved in nuclear explosions (e.g., [105]). Two groups were chosen to develop the simulation capability: one at the Naval Research Laboratory (NRL) and the other at Mission Research Corporation (MRC). Both groups developed global simulation models for the ionosphere and magnetosphere using advanced numerical methods [106, 107]. These models became known as the Lyon-Fedder-Mobarry (LFM) [6, 108] and Integrated Space Weather Model (ISM) [109].

Of the early global magnetospheric MHD codes developed at UCLA in the 1970s, only the Ogino code [24, 104, 110] survived; the others were not developed beyond the proof-of-concept stage. A decade later, two more codes were developed in Japan: the relatively simple Watanabe-Sato code [9, 111] and the advanced Tanaka model [112, 113].

In the 1990s, NASA, ESA, and ISAS of Japan jointly undertook the International Solar-Terrestrial Physics (ISTP) Science Initiative. In the framework of this initiative, NASA supported the transition of the LFM code from NRL to the University of Iowa and the development of a new 3D global MHD code at UCLA, which later became known as OpenGGCM [7, 114].

At about the same time, three other global MHD models were developed describing the magnetosphere. Robert Winglee at the University of Washington further developed Steinolfson's model and applied it to the magnetosphere [26, 115], Pekka Janhunen of the Finnish Meteorological Institute (FMI) developed a 3D adaptive mesh refinement (AMR) code using a relatively simple first-order Godunov solver [11, 116], and Gombosi and coworkers applied modern computational fluid dynamics (CFD) methods to develop the 3D AMR global MHD code called BATS-R-US (Block-Adaptive Tree Solar-Wind Roe-Type Upwind Scheme) [29, 117]. The development of the eight-wave solver [117] at the heart of early BATS-R-US created some controversy [118–120], but it eventually died out with increasing evidence in support of the validity of the Michigan magnetosphere model.

In the early 2000s, the Space Weather Modeling Framework (SWMF) was developed [30, 31]. It provides a software environment and tools for coupling various models with each other. It is a fully functional, documented software framework that allows parallel execution and efficient coupling of multiple models [121]. Presently SWMF is the only operational software framework available for the geospace community that enables high-performance model coupling.

Table 7.1 lists the global magnetosphere models with their main characteristics. We note that LFM, GUMICS, OpenGGCM, and SWMF/BATS-R-US are available for community use through the Community Coordinated Modeling Center located at NASA Goddard Space Flight Center [124].

Geospace models at the CCMC. The Community Coordinated Modeling Center (CCMC) is a multi-agency partnership located at NASA Goddard Space Flight Center [124]. While NASA provides most of the funding for CCMC, other

Table 7.1 List of global 3D MHD geospace models

Code	Institution	Method	I-M coupling	Ring current	References
Ogino	U. Nagoya	Lax-Wendroff + UG	FAC + resistive sphere	No	[24, 104, 110]
LFM	Dartmouth	FV-TVD + SG	FAC + Φ_{ionos}	Yes	[6, 108]
ISM	MRC	FV-TVD + SG	FAC + Φ_{ionos}	No	[109]
Watanabe-Sato	Hiroshima U.	Lax-Wendroff + PG	FAC + resistive sphere	No	[9, 111]
OpenGGCM	UNH/UCLA	FV-TVD + SG	FAC + Φ_{ionos}	Yes	[7, 114]
SWMF/BATS-R-US	U. Michigan	FV-TVD + SG + AMR	FAC + Φ_{ionos}	Yes	[29–31]
GUMICS	FMI	Godunov + AMR	FAC + Φ_{ionos}	No	[11, 116]
Winglee	U. Washington	Lax-Wendroff + PG	FAC + resistive sphere	No	[26, 115]
Tanaka	Kyushu U.	FV-TVD + UnG	FAC + Φ_{ionos}	No	[112, 113]
PPMLR-MHD	USTC/CSSAR	FV-TVD + SG	FAC + Φ_{ionos}	No	[122, 123]
GAMERA	APL	FV-TVD + SG	FAC + Φ_{ionos}	Yes	[86]

Abbreviations: FV-TVD, finite-volume total variation diminishing; AMR, adaptive mesh refinement; UG, uniform grid; SG, stretched grid; PG, patched grid; UnG, unstructured grid; Φ_{ionos} , electrostatic ionospheric potential; FAC, field-aligned currents

federal agencies also support its mission. These agencies include the Air Force Office of Scientific Research (AFOSR), the Air Force Research Laboratory (AFRL), the Air Force Weather Agency (AFWA), the National Science Foundation (NSF), and the NOAA Space Weather Prediction Center (SWPC).

BATS-R-US [29] was the first model transitioned to CCMC (in 2000) to describe the global magnetosphere. In 2001, BATS-R-US became the first model that was available for Runs-on-Request (RoR). It has also been running in real-time mode at CCMC since 2002. In 2004, the Space Weather Modeling Framework (SWMF) [31] was transitioned to CCMC. The SWMF enabled the coupling of global and regional models, and it greatly enhanced the modeling resources available for the space physics community.

Presently CCMC hosts a large number of codes describing solar, heliospheric, magnetospheric, ionospheric, and thermospheric phenomena. These include the following physics-based global geospace models:

- The Lyon-Fedder-Mobarry (LFM) global MHD model [6] and its coupling to ionospheric electrodynamics [125] and ionosphere-thermosphere models [126–128]. LFM is a time-dependent, ideal MHD magnetosphere model. It uses a TVD transport scheme [129] with eighth-order reconstruction in smooth regions. LFM uses the constrained transport method [107] to maintain the $\nabla \cdot \mathbf{B} = 0$ condition to roundoff. The LFM grid is a distorted spherical grid with azimuthal symmetry about the polar axis. It usually covers the domain from about $30 R_E$ upwind to $300 R_E$ downwind of the Earth and roughly $100 R_E$ out to the sides.
- The Open Geospace General Circulation Model (OpenGGCM) was developed at UCLA in the early 1990s [8]. It solves the resistive MHD equations in the magnetosphere with an ionospheric boundary condition consisting of an ionosphere potential solver that closes the magnetospheric field-aligned currents (FAC) in the ionosphere. Later the OpenGGCM was extended with a coupling module that maps the FAC into the ionosphere and the potential back into the magnetosphere and also computes electron precipitation parameters, as well as the ionospheric Hall and Pedersen conductances using empirical relations [114]. OpenGGCM solves the resistive MHD equations on a stretched Cartesian grid using second-order explicit time integration with conservative and flux-limited spatial finite differences. Constrained transport [107] is used to preserve the magnetic field divergence to roundoff errors.
- The Grand Unified Magnetosphere-Ionosphere Coupling Simulation (GUMICS) code is a global solar wind-magnetosphere-ionosphere coupling model [11, 116, 130]. Its solar wind and magnetospheric part is based on solving the ideal MHD equations, and its ionosphere part is based on solving the electrostatic current continuity equation in a high-resolved upper atmosphere model. The latest version (GUMICS-4) uses a first-order Godunov scheme with the eight-wave scheme [29] and cell-based adaptive Cartesian grid refinement [116] with temporal subcycling. The computational box is from -224 to $+32$ Earth radii in X (pointing from the Earth toward the Sun) and from -64 to $+64$ Earth radii in the other two directions.

- The geospace modeling configuration [131] of the Space Weather Modeling Framework [30,31] (SWMF/Geospace) couples together several domain models: (i) the Block-Adaptive Tree Solar-Wind Roe-Type Upwind Scheme (BATS-R-US) [29] that solves 3D MHD equations in finite-volume form using numerical methods related to Roe's approximate Riemann solver (it uses an adaptive grid composed of rectangular blocks arranged in varying degrees of spatial refinement level); (ii) an ionospheric potential solver that provides electric potentials and conductances in the ionosphere from magnetospheric FAC [132]; (iii) and one of the available inner magnetosphere (ring-current) models [133–135].

7.3 The Space Weather Modeling Framework (SWMF) [30, 31]

In this paper, we describe the SWMF and its available geospace components as an example of adaptive global simulation tools of the near-Earth space environment. It represents the state of the art in large-scale space physics simulations, and it demonstrates the challenges facing modern space plasma simulation codes.

SWMF is [30,31] a fully functional, documented technology (the user manual is over 200 pages long) that provides a high-performance computational capability to simulate the space-weather environment from the upper solar chromosphere to the Earth's upper atmosphere and/or the outer heliosphere. The SWMF is currently used by several universities and research laboratories, including the CCMC at NASA Goddard Space Flight Center [124]. The source code is publicly available at <http://csem.engin.umich.edu>, and SWMF runs can be requested via the CCMC.

Figure 7.4 shows the present structure of the SWMF. There are over a dozen components or physics domains. The arrows show how the domains are coupled together. In an actual simulation, one can use any meaningful subset of the components. If the simulation starts from the Sun, it is typically driven by synoptic magnetogram data and flare/CME observations. Simulations restricted to magnetospheric components are usually driven by the solar wind data obtained by satellites upstream of the Earth, for example, ACE, Wind, or Geotail. We also use the F10.7 solar flux for some of the empirical relationships in the ionosphere and thermosphere models.

The SWMF has a layered architecture (see Fig. 7.5). The top layer is an optional graphical user interface. The second layer contains the control module, which is responsible for distributing the active components over the parallel machine, executing the models, and coupling them at the specified frequencies. The third layer contains the physics domain components. Each component can have multiple physics models. Each component version consists of a physics model with a wrapper and one or more couplers. The wrapper is an interface with the control module, while each coupler is an interface with another component. The physics models can also be compiled into stand-alone executables. The fourth and lowest layer contains the shared library and the utilities that can be used by the physics models as well as by the SWMF core.

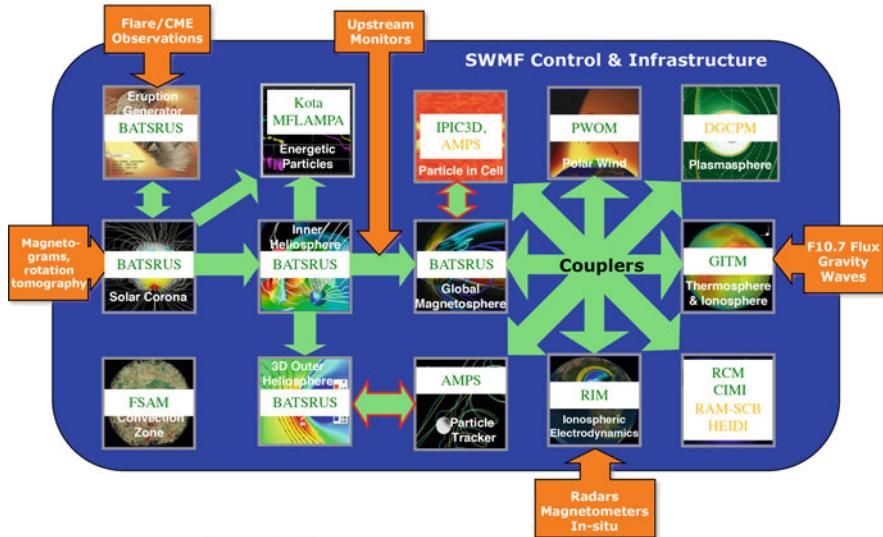


Fig. 7.4 Schematic diagram of the Space Weather Modeling Framework. In practice, the SWMF is almost never run with all components coupled together, but we typically use a subset of the available components. For example, we can run the solar corona and the inner heliosphere (AWSoM) models driven by solar synoptic magnetograms [94] or the global magnetosphere, inner magnetosphere, and ionosphere electrodynamics models (SWMF/Geospace) driven by L1 satellite observations (reproduced from [131] with the permissions of AGU)

SWMF couples together the various models at regular intervals, based on either simulation time or iteration number. The relevant physical quantities are passed with efficient MPI communication. In addition to transferring the data, SWMF has to transform between coordinate systems, take care of unit conversions, and interpolate between different grids. Often the models are moving or rotating relative to each other so that the mapping has to be recalculated every coupling time. A further complication arises for adaptive grids that may change between two couplings. SWMF includes utilities to take care of coordinate transformations and interpolation between various grids.

Since the models use widely different grids and time-steps, coupling through a simple interface may be very challenging, especially when the flow is slower than the fast magnetosonic speed. A possible solution is to overlap the models. For example, the inner boundary of the inner heliosphere model is provided by the solar corona model at 20 solar radii, while solar corona obtains its outer boundary conditions from inner heliosphere module at 24 solar radii. The overlap serves as a buffer to suppress numerical artifacts due to the differences between the spatial and temporal resolutions.

In some cases, the coupling between the physics models requires some complicated and expensive calculations. For example, the inner magnetosphere and the radiation belt models require the magnetic field geometry and the plasma state

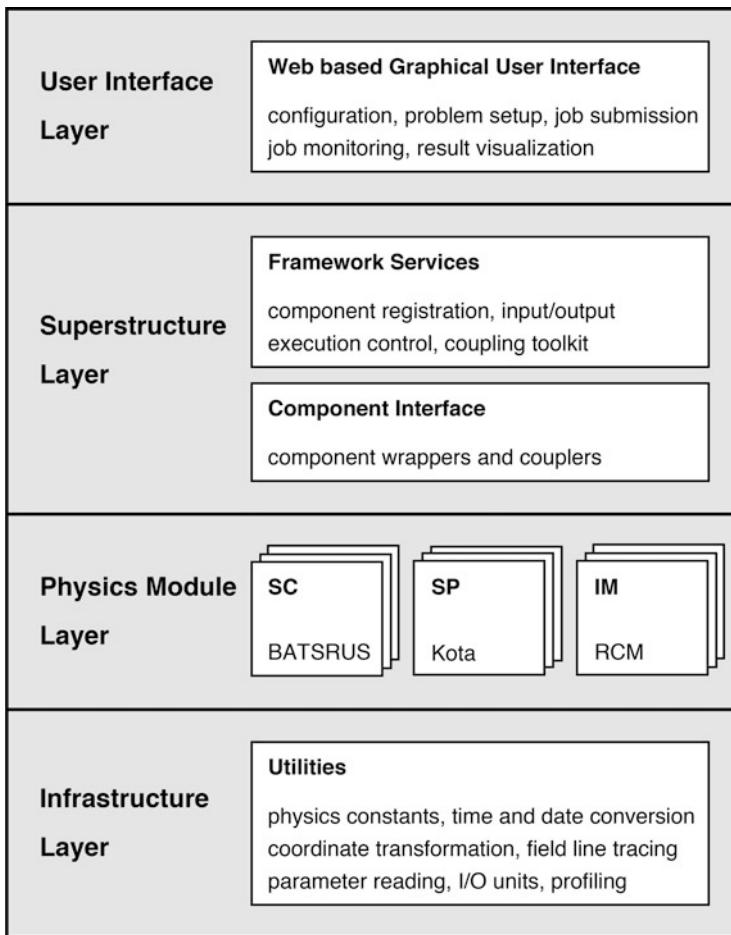


Fig. 7.5 The layered architecture of the SWMF (reproduced from [30] with the permission of Elsevier)

along the closed magnetic field lines of the global magnetosphere model. Since the global magnetosphere grid is large and distributed over many processors, the tracing of magnetic field lines is quite challenging. SWMF uses a couple of highly parallel and efficient schemes for tracing multiple field lines [136, 137] that provide mapping information, integrate quantities along the lines, or extract state variables and positions along the lines.

7.4 BATS-R-US

The BATS-R-US code [29, 30] plays a central role in the SWMF. It is a versatile, high-performance, generalized magnetohydrodynamic code with adaptive mesh refinement (AMR) that can be configured to solve the governing equations of ideal and resistive MHD [29]; semi-relativistic [138], anisotropic [139], Hall [140], multispecies [141], and multifluid [142] extended magnetofluid equations (XMHD); and most recently non-neutral multifluid plasmas [143]. BATS-R-US is used to model several physics domains (see Fig. 7.4). The efficiency of BATS-R-US is crucial to reach faster than real-time performance with the SWMF while maintaining high resolution in the domains of interest.

7.4.1 Modular Architecture

BATS-R-US has a layered modular software architecture to handle several applications with a single base code. The state variables of the equation system are defined by the equation modules, while the rest of the application-dependent details are implemented into user modules. A configuration script is used to select the equation and user modules that are compiled together with the code. There are currently dozens of equation and user modules (obviously not all combinations are possible) which means that BATS-R-US can be configured for quite a few different applications. Time stepping is also very flexible as can be seen in Fig. 7.6. In addition to the basic equations, there are various source terms that change from application to application: collisions, charge exchange, chemistry, photo-ionization, recombination, radiative losses, etc. The boundary and initial conditions vary greatly as well. The major features of BATS-R-US are summarized in Fig. 7.9.

7.4.2 Block-Adaptive Mesh Refinement

BATS-R-US uses a simple and effective block-based adaptive mesh refinement (AMR) technique [29]. The governing equations are integrated to obtain volume-averaged solution quantities within computational cells. The computational cells are embedded in regular structured blocks of equal-sized cells. The blocks are geometrically self-similar. Typically, the blocks we use consist of anywhere between $4 \times 4 \times 4 = 64$ and $12 \times 12 \times 12 = 1728$ cells (see Fig. 7.7). Solution data associated with each block are stored in standard indexed array data structures, making it 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 can be of different sizes in terms of the volume of physical space they occupy. Starting with an initial mesh consisting of blocks of equal size (i.e., uniform resolution), spatial

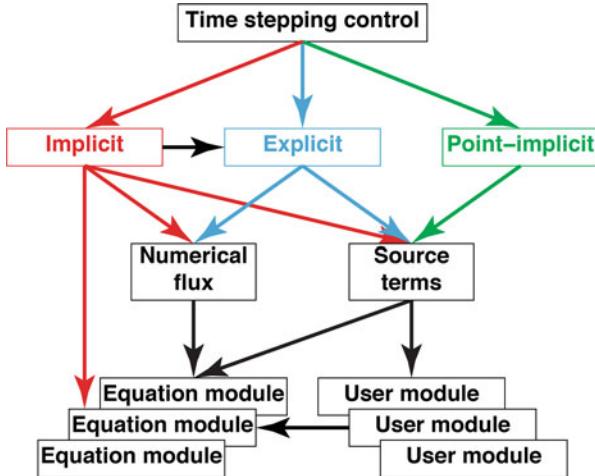


Fig. 7.6 The layered software structure of BATS-R-US. The arrows point from the module that is using data or methods from the other module. There are multiple versions of the equation and user modules containing the equation variable definitions and the application-specific code, respectively. The numerical fluxes depend on the selected equation module, and the source terms may also be defined in the user module. The various time-stepping schemes are independent of the details of the equations being solved with the possible exception of the implicit preconditioner (reproduced from [30] with the permission of Elsevier)

adaptation is performed by dividing and coarsening 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 with the same number of cells as the parent, which doubles cell resolution in the region of interest. Conversely, in overresolved regions, the refinement process reverses; eight children coarsen and coalesce into a single parent block. Thus, cell resolution reduces by a factor of 2. Multigrid-type restriction and prolongation operators are used to evaluate the solution on all blocks created by the coarsening and division processes, respectively.

When a 3D block is refined, it is split into eight octants (see Fig. 7.7). Each octant forms a block with the same number of cells as the original block, but the resolution is increased by a factor of 2. The resulting grid structure is an octree of blocks, and the equations are solved at the finest level only, i.e., on the leaves of the tree (see Fig. 7.8).

The hierarchical data structure and self-similar blocks simplify domain decomposition and enable good load-balancing, a crucial element for truly scalable computing. For explicit time-stepping (all blocks use the same time-step), natural load-balancing occurs by distributing the blocks equally among the processors. For more complicated time-stepping schemes, the load-balancing is more challenging, but it still can be done on a block-by-block basis. We achieve additional optimization by ordering the blocks using the Peano-Hilbert space-filling curve to minimize inter-

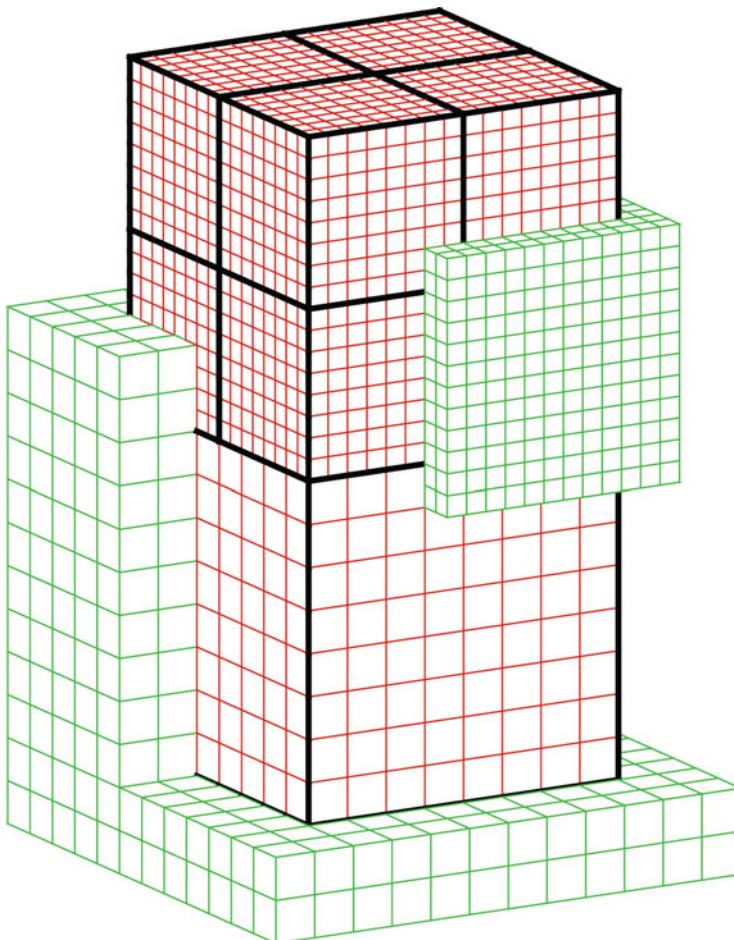


Fig. 7.7 Self-similar blocks illustrating the double layer of ghost cells for both coarse and fine blocks (reproduced from [144] with the permission of Springer publishers)

processor communication. The self-similar nature of the solution blocks also means that serial performance enhancements apply to all blocks and that fine-grained algorithm parallelization is possible. The algorithm's parallel implementation is so pervasive that even the grid adaptation performs in parallel.

7.4.3 Conservation Laws

BATS-R-US can solve many approximations to the low-order velocity moments of the Boltzmann equations (we refer the interested readers to [145–149]. The

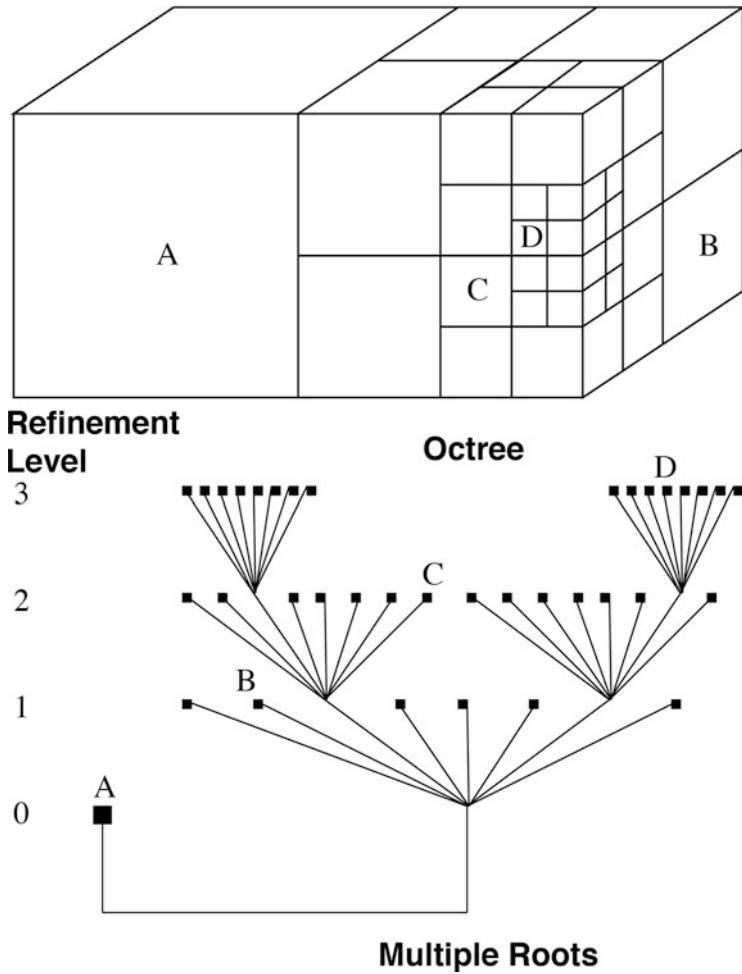


Fig. 7.8 Solution blocks of the BATS-R-US computational mesh with three refinement levels (reproduced from [144] with the permission of Springer publishers)

governing equations for species “s” can be written as

$$\frac{\partial \rho_s}{\partial t} + \nabla \cdot (\rho_s \mathbf{u}_s) = 0 \quad (7.1a)$$

$$\frac{\partial \rho_s \mathbf{u}_s}{\partial t} + \nabla \cdot [\rho_s \mathbf{u}_s \mathbf{u}_s + p_{s\perp} \mathbf{I} + (p_{s\parallel} - p_{s\perp}) \mathbf{b} \mathbf{b}] = \frac{q_s}{m_s} \rho_s (\mathbf{E} + \mathbf{u}_s \times \mathbf{B}) \quad (7.1b)$$

$$\frac{\partial p_{s\parallel}}{\partial t} + \nabla \cdot (p_{s\parallel} \mathbf{u}_s) = -2p_{s\parallel} \mathbf{b} \cdot (\mathbf{b} \cdot \nabla) \mathbf{u}_s \quad (7.1c)$$

$$\frac{\partial p_{s\perp}}{\partial t} + \nabla \cdot (p_{s\perp} \mathbf{u}_s) = -p_{s\perp} (\nabla \cdot \mathbf{u}_s) + p_{s\perp} \mathbf{b} \cdot (\mathbf{b} \cdot \nabla) \mathbf{u}_s \quad (7.1d)$$

where ρ and \mathbf{u} denote the mass density and the velocity vector, respectively, and q and m are the charges and masses of the particles. For the pressure tensor, we used the CGL approximation [150], $\mathbf{P} = p_{\perp}\mathbf{I} + (p_{\parallel} - p_{\perp})\mathbf{bb}$, where \mathbf{I} is the identity matrix, \mathbf{b} is the unit vector along the magnetic field direction, p_{\parallel} is the pressure along the parallel direction of the magnetic field, and p_{\perp} is the pressure in the perpendicular direction. The scalar pressure can be written as $p = (p_{\parallel} + 2p_{\perp})/3$. BATS-R-US has the capability to solve the full equation system (7.1) or reduce it and only solve for the scalar pressure, p .

The electric (**E**) and magnetic fields (**B**) are obtained from Maxwell's equations:

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} = 0 \quad (7.2a)$$

$$\frac{\partial \mathbf{E}}{\partial t} - c^2 \nabla \times \mathbf{B} = -c^2 \mu_0 \mathbf{j} \quad (7.2b)$$

$$\nabla \cdot \mathbf{E} = \frac{\rho_c}{\epsilon_0} \quad (7.2c)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (7.2d)$$

where ϵ_0 is the vacuum permittivity, μ_0 is the vacuum permeability, $c = 1/\sqrt{\epsilon_0 \mu_0}$ is the speed of light, $\rho_c = \sum_s (q_s/m_s) \rho_s$ is the total charge density, and $\mathbf{j} = \sum_s (q_s/m_s) \rho_s \mathbf{u}_s$ is the current density. Equations 7.2c and 7.2d are constraints on the initial conditions, and analytically these conditions are preserved. Numerically, however, this is not guaranteed to hold. BATS-R-US uses a variety of methods to enforce the solenoidal magnetic field condition (for more details, see [30]).

Extended magnetohydrodynamics (XMHD) makes two fundamentally important assumptions: (i) electrons are assumed to be massless and (ii) charge neutrality is assumed at all scales. These two assumptions lead to the generalized Ohm's law:

$$\mathbf{E} = -\mathbf{u}_e \times \mathbf{B} - \frac{1}{en_e} \nabla \cdot [p_{e\perp} \mathbf{I} + (p_{e\parallel} - p_{e\perp}) \mathbf{bb}] \quad (7.3)$$

In a single-ion plasma, the electron velocity is $\mathbf{u}_e = \mathbf{u}_i - \mathbf{j}/(en_e)$ resulting in the motional electric field plus the Hall term. The second term in Eq. 7.3 is the ambipolar electric field. It is interesting to note that the parallel (field-aligned) component of the electric field is

$$\mathbf{E}_{\parallel} = \mathbf{b} \cdot \mathbf{E} = -\frac{\nabla_{\parallel} p_{e\parallel}}{en_e} + \frac{p_{e\parallel} - p_{e\perp}}{en_e} \frac{\nabla_{\parallel} B}{B} \quad (7.4)$$

where $\nabla_{\parallel} = \mathbf{b} \cdot \nabla$ is the parallel gradient operator. In Eq. 7.4, the first term describes the parallel ambipolar electric field, while the second term represents adiabatic focusing. BATS-R-US has the capability to solve various XMHD approximations, from ideal MHD to resistive, Hall, anisotropic pressure and multispecies and

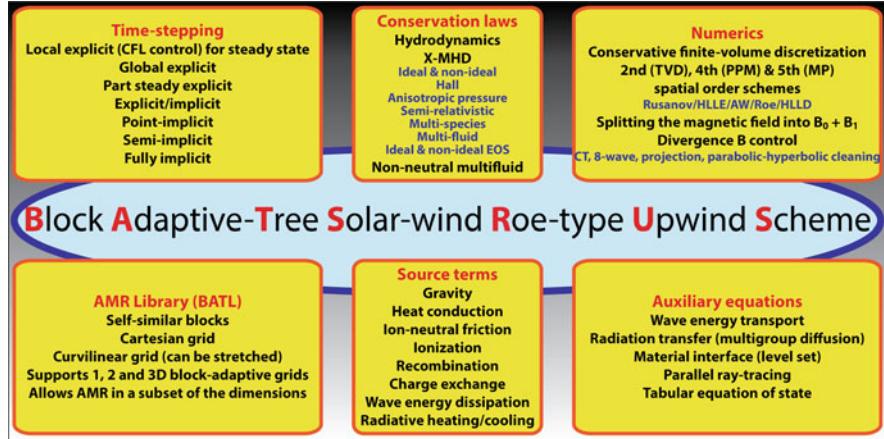


Fig. 7.9 Overview of the BATS-R-US multiphysics code (reproduced from [30] with the permission of Elsevier)

multifluid limits. A more detailed description of these capabilities can be found in [30].

A recent addition to the BATS-R-US equation set is the six-moment approximation [143]. This approximation solves the full set of Eqs. 7.1 and 7.2 without neglecting the electron mass and assuming charge neutrality. Consequently, there is no Ohm's law to express the electric field, and we need to solve the full set of electron equations. This approximation introduces new wave modes, including the light wave (with a characteristic speed of the speed of light) and the whistler wave. Both of these waves have speeds well exceeding the typical MHD wave speeds. These new waves represent additional challenges for the numerical solution, but they introduce new physical wave modes that make the solutions more realistic. Details of this method were recently published by Huang et al. [143]. A concise summary of the BATS-R-US code is shown in Fig. 7.9.

7.4.4 BATS-R-US Performance

Large massively parallel distributed-memory machines offer huge CPU and memory resources. Capitalizing on these resources is, however, not always easily achieved. For most computational models that involve the solution of PDEs, domain decomposition (i.e., partitioning the problem by dividing the computational domain into subdomains and farming the subdomains off onto separate cores) is a natural and, in many cases, the most practical approach to parallelization. In many instances, solver speedup is achieved for a modest number of cores; however, with added cores, not only does the method fail to scale as expected but also the performance of the

algorithm may actually diminish with an increase in the number of cores. These inherent limitations can arise for a number of reasons including:

1. The underlying solution algorithm is global in nature and requires high levels of inter-processor communication.
2. The underlying data structure is not readily partitioned and/or updated in a parallel fashion.
3. The underlying physical model and mathematical description is inherently serial.

In order to avoid many of the limitations described above, the 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 has a compact stencil and is therefore highly local in nature. This results in low inter-processor communication overhead. It also permits the more efficient use of memory and cache. 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 among the processors, and for ten blocks per processor, the load imbalance is less than 10% (the load imbalance is less than 1% for 100 blocks per node). 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 for programming language and the message passing interface (MPI) library for performing the message passing. Use of these standard programming tools 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 such that it typically accounts for less than 3–5% of processor time. The serial performance of the algorithm, and hence the overall parallel performance of the method, has also been greatly enhanced by avoiding the use of indirect addressing and allocatable arrays when defining memory for primary solution variables, by more efficient use of cache, and by performing loop optimizations. Figure 7.10 shows the weak scaling (how the solution time varies with the number of processors for a fixed problem size per processor) of BATS-R-US on several machines. We note that these scaling runs were carried out a few years ago and today the code performance is even better.

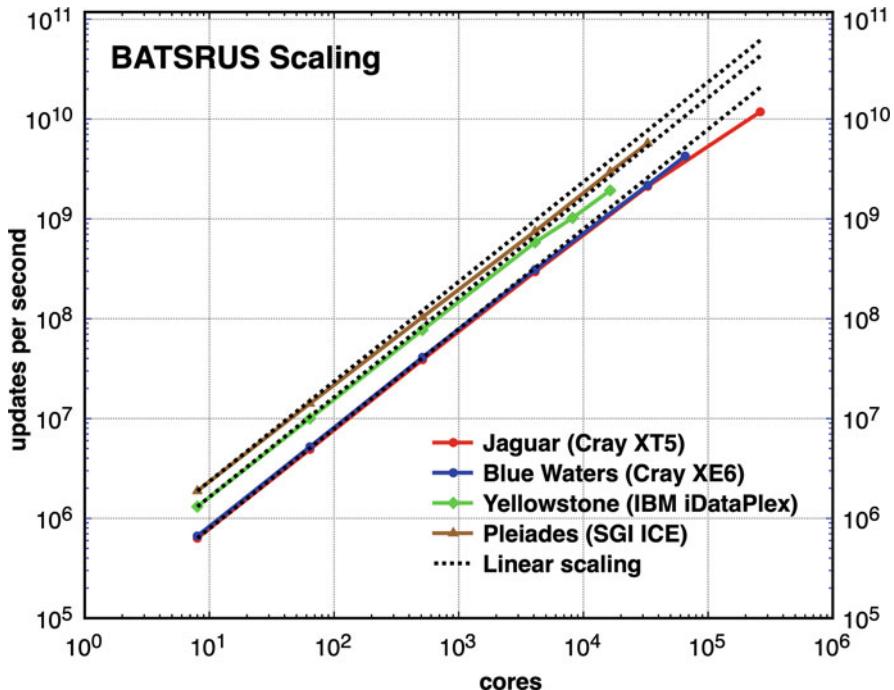


Fig. 7.10 The cell update rate as a function of number of cores for BAT-S-R-US. The problem size scales in proportion to the number of parallel processes

7.5 Heliophysics and Planetary Applications

Over the last couple of decades, the SWMF/BATS-R-US simulation capability has been extensively used to study the solar corona (e.g., [94, 95, 151–160]), solar wind (e.g., [47, 92, 153, 161–166]), coronal mass ejections (e.g., [167–180]), and solar energetic particles (e.g., [181–186]). It was also used to describe the interaction of the solar wind with the interstellar medium (e.g., [187–192]). With the efficient coupling, it was also used in simulating the entire Sun-Earth system (e.g., [193–197]) as well as the complex terrestrial magnetosphere-ionosphere system (e.g., [136, 142, 198–204]) and space weather (e.g., [131, 205–208]).

SWMF/BATS-R-US is also used to model planetary plasma environments. It has been used to simulate the magnetospheres of Mercury (e.g., [209–212]), Venus (e.g., [213, 214]), Mars (e.g., [141, 215–219]), Jupiter (e.g., [220, 221]), Saturn (e.g., [222–229]), Uranus [230], comets (e.g., [231–238]), as well as planetary satellites (Io [239, 240, 240], Europa [241–243], Ganymede [244, 245], Titan [246–249], and Enceladus [250–252]). In addition, it has been used by the DoE-supported Center for

Radiative Shock Hydrodynamics to simulate laser plasma interactions (e.g., [253–255]).

Here we present an example of SWMF/BATS-R-US simulations of the geospace environment [131]. Such runs are available for the space science community in several places. First, they can be accessed at the CCMC through their experimental real-time simulation webpage (<https://ccmc.gsfc.nasa.gov/cgi-bin/SWMFpred.cgi?run=SWMF-RCM>). This is the most versatile of the options because it enables the user to use the standard CCMC plotting software to explore the full range of output parameters and display styles from each of the codes within the SWMF suite (i.e., SWMF/Geospace comprised of BATS-R-US, RCM, and RIM). Output is also available through CCMC’s Integrated Space Weather Analysis (iSWA) website (<https://ccmc.gsfc.nasa.gov/iswa/>). Two examples of the output available at iSWA are shown in Fig. 7.11, displaying 2D output slices from BATS-R-US during a relatively quiet time with nominal solar wind input. The main features seen in the panels of Fig. 7.11 are the subsolar magnetopause near $10 R_E$ and the inner edge of the nightside plasma sheet near geosynchronous distance. Yet another place where output is available is a dedicated page at the University of Michigan (U-M) Center for Space Environment Modeling (CSEM) (<http://csem.engin.umich.edu/realtime/index.php>). The CSEM site allows users to see the latest images of the full set of iSWA plots of the SWMF output, from both experimental real-time versions (i.e., with and without RCM). This site also shows a comparison of the model values for Dst displayed against the observed real-time Dst values from the Kyoto World Data Center (WDC) for geomagnetism. Figure 7.12 shows an example set of these Dst comparisons for 3 months in 2015. This interval was chosen because it contains a relatively large number of moderate storms, showcasing the ability of SWMF to capture the time series of measured geomagnetic activity.

One can see in Fig. 7.12 that SWMF/Geospace can describe space weather effects on the near-Earth current system as indicated by the capture of surface magnetic fluctuations and virtual magnetometer results (see [256, 257]). Detailed validation studies can also be found in [131, 208, 258].

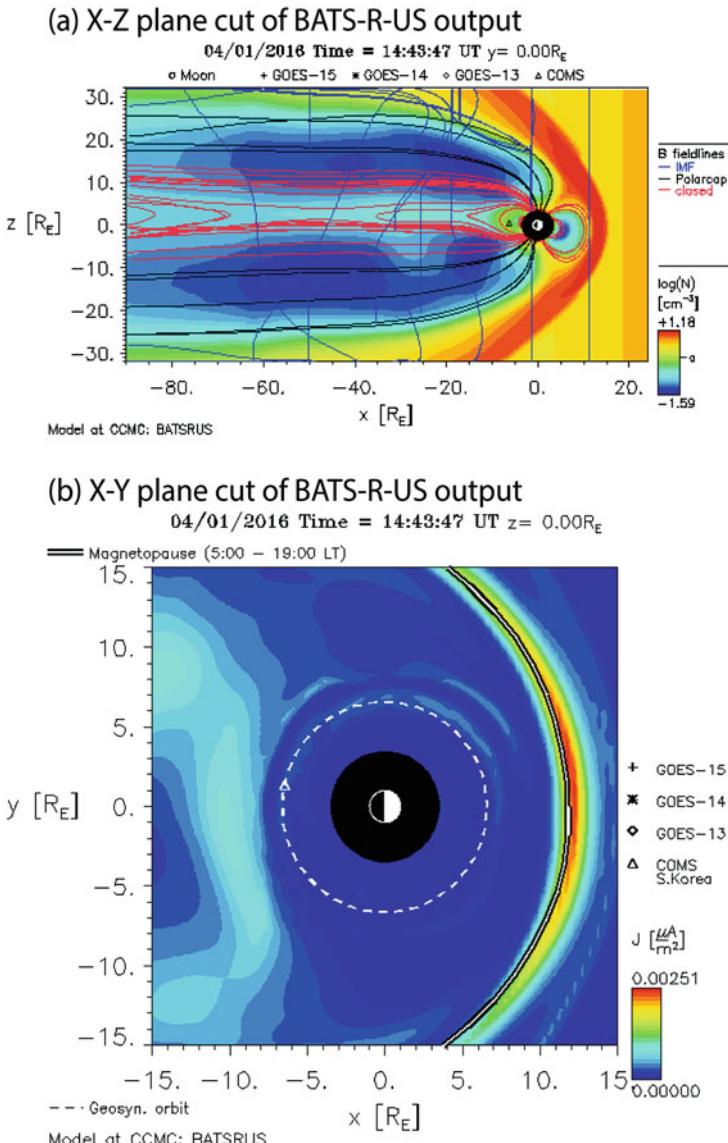


Fig. 7.11 Example of output from the experimental real-time SWMF/Geospace simulations at CCMC, showing two plots from the BATS-R-US code during a relatively quiet interval with near-nominal solar wind values: (a) number density in the meridional (X-Z) plane and (b) magnitude of current density in the equatorial (X-Y) plane. The overlaid lines in panel (a) show the magnetic field color coded by their connectivity topology (blue, IMF; black, open; red, closed). The white dashed circle in panel (b) shows the location of geosynchronous orbit (reproduced from [131] with the permission of AGU)

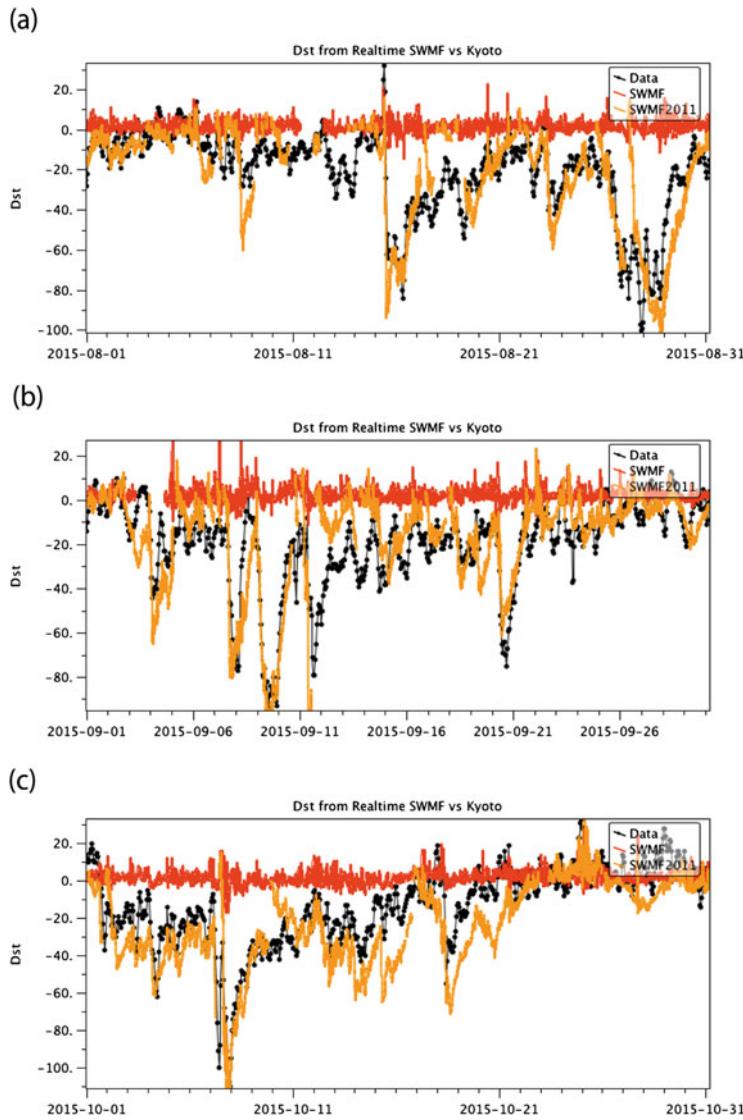


Fig. 7.12 Example of output from the CCMC experimental real-time runs of SWMF/Geospace, showing Dst comparisons for 3 months (August–September 2015). The black line with dots is the real-time Dst time series from the Kyoto WDC, the yellow line is the SWMF result with the RCM inner magnetosphere model included, and the red line is the earlier version of the experimental real-time code without the RCM (reproduced from [131] with the permission of AGU). **(a)** SWMF-Geospace Dst comparison for August 2015. **(b)** SWMF-Geospace Dst comparison for September 2015. **(c)** SWMF-Geospace Dst comparison for October 2015

7.6 MHD-EPIC

A new modeling capability to embed the implicit particle-in-cell (PIC) model iPIC3D [259] into the BATS-R-US magnetohydrodynamic model [140] has been developed to investigate small-scale kinetic effects in global XMHD simulations [36, 260, 261]. The PIC code has recently been improved by including a new energy conserving algorithm [262] and a $\nabla \cdot \mathbf{E}$ correction [263].

The PIC domain can cover the regions where kinetic effects are most important, such as reconnection sites. The BATS-R-US code, on the other hand, can efficiently handle the rest of the computational domain where the MHD or Hall MHD description is sufficient. Since the PIC model is able to describe self-consistently the electron behavior, our coupled MHD-EPIC model is well suited for investigating the nature of magnetic reconnection in reconnection-driven magnetospheres.

7.6.1 iPIC3D

iPIC3D is a parallel high-performance implicit particle-in-cell (PIC) code [259]. It solves the full set of Maxwell's equations for the electromagnetic fields coupled with the equations of motion for electrons and ions on 3D Cartesian grids. The discretization is based on the implicit moment PIC (IMPiC) method that employs an implicit time integration for the electric field, then the magnetic field is updated from the induction equation, and finally the particles are moved with a simple iterative scheme [264–267]. The main advantage of iPIC3D is that it is capable to take larger grid cell size and time-steps and thus make the coupled simulation affordable on today's supercomputers.

A companion chapter in this book [268] discusses in detail the challenges and compromises explicit and implicit PiC codes make. The bottom line is that if one wants to resolve Debye scale phenomena, the use of expensive explicit PiC codes is necessary. However, if one is mainly interested in reconnection and other space plasma phenomena, the use of implicit PiC codes is not only justified but also necessary (cf. [268, 269]).

7.6.2 Coupling MHD and PiC Simulations

Figure 7.13 illustrates the overall flow of the coupling algorithm between BATS-R-US and iPIC3D [36], while Fig. 7.14 shows the spatial discretization of the coupling. It is important to point out that the BATS-R-US-iPIC3D coupling via SWMF is genuinely two-way: all physical quantities are self-consistently advanced by both codes and the relevant information is fully exchanged in every time-step. This coupling is different from the one-way coupled approach developed by the UCLA group [35, 270, 271].

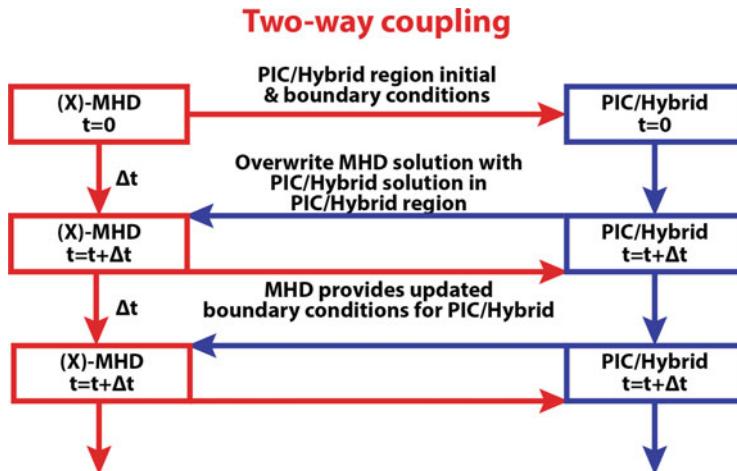


Fig. 7.13 The overall flow of MHD-PIC coupling. At $t = 0$, the MHD code sends the description of the PIC region (size and grid resolution) and the MHD state inside and around the PIC region to the PIC code, so it can set up the PIC domain and initial state. Both the MHD and PIC codes then advance by one or more time-steps until both models reach the next coupling time. Again, information is exchanged both ways, but this time the PIC code only uses the MHD solution as a boundary condition, while the MHD code overwrites the PIC region with the PIC solution. Both codes continue to advance to the next coupling time, exchange information, and so on, until they reach the final simulation time or until the PIC region is removed (after [36] with the permission of Elsevier)

MHD-EPIC [36, 245, 260, 261] uses the node-centered number densities, velocities, pressures, and magnetic field (large red dots in Fig. 7.14) to create the macroparticles inside the ghost cells of the PIC grid as illustrated by the small red dots in the light gray area in the figure. The particles that leave the PIC region (dark gray area in the figure) are discarded at the end of the PIC time-step. New macroparticles are generated for each species in each ghost cell of the PIC domain with the appropriate (bi-)Maxwellian distribution functions using the MHD solution. The locations of the new particles are random with a uniform distribution over the ghost cell. For each ghost cell, the corresponding number density, velocity, and pressure are linearly interpolated from the surrounding MHD values (large red dots in the figure) to the given location.

In this two-way coupled method, the MHD values in the cell centers covered by the nodes of the PIC grid (black squares in Fig. 7.14) are fully overwritten by the PIC solution. The magnetic field can simply be interpolated from the PIC field. For the other MHD variables, MHD-EPIC [36, 245, 260, 261] takes various moments of the distribution function represented by the macroparticles. Since iPIC3D already contains the algorithms that calculate the zeroth, first, and second moments of the distribution function at the cell nodes, one can take full advantage of these to obtain the MHD values.

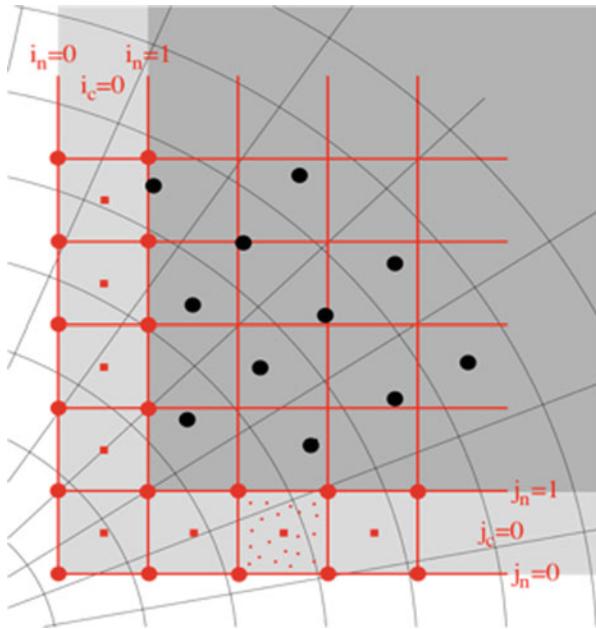


Fig. 7.14 Spatial discretization of the MHD-EPIC coupling. The Cartesian grid of the PIC region is indicated with the darker gray area surrounded by the thick red line. The lighter gray area shows the ghost cell/node region of the PIC grid. The large red dots are node values fully or partially obtained from the MHD variables. The small red dots illustrate particles created in the ghost cells of the PIC grid. The small red squares are the ghost cell centers of the PIC grid where the magnetic field is set from the MHD solution. The black squares indicate the MHD cell centers where the solution is obtained from the PIC code. The MHD grid can be either Cartesian or spherical (after [36] with the permission of Elsevier)

7.6.3 MHD-EPIC Performance

The SWMF and the BATS-R-US codes require a Fortran compiler (the source code is written mostly in F90 with a few F95 and F2003 features) and the MPI library. The iPIC3D code requires a C++ compiler and the MPI library and optionally the parallel HDF5 library. Very good scaling up to 32k MPI processes on 1024 XE nodes of the Blue Waters supercomputer were achieved as shown in Figs. 7.10 and 7.15.

Currently SWMF, BATS-R-US, and iPIC3D mainly use pure MPI parallelism that works fine up to 1024 XE nodes and 32k MPI processes (see Figs. 7.10 and 7.15). While this is more than sufficient for most applications, the codes encounter some limitations when running with 65k and more MPI processes. Efforts are underway to use OpenMP parallelism on the nodes to reduce the amount of memory usage and the number of messages sent between the MPI processes [245]. The OpenMP+MPI approach does not improve the performance relative to the pure MPI parallelization, but the hybrid approach allows running larger problems on larger number of nodes [245].

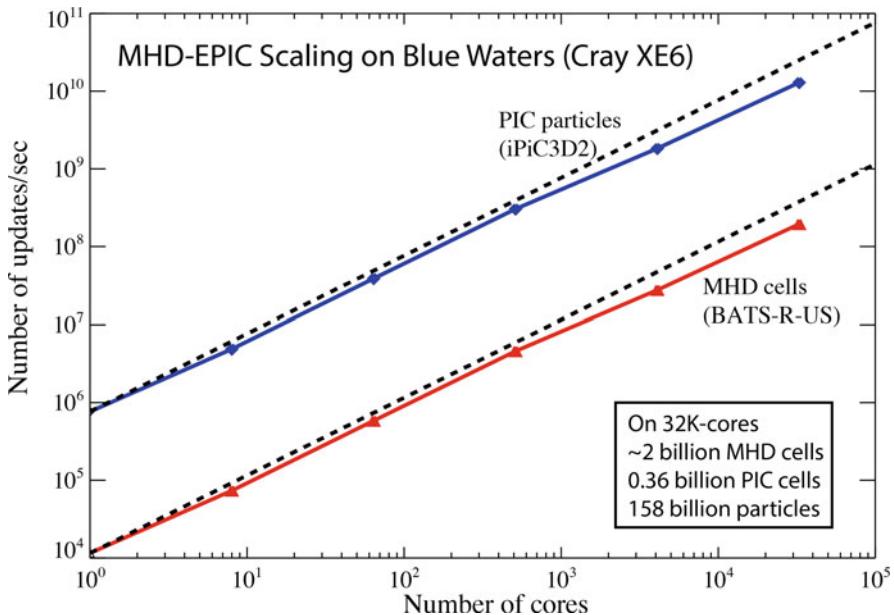


Fig. 7.15 MHD-EPIC scaling on Blue Waters up to 32 k MPI processes on 1024 nodes. The blue curve shows the number of particle updates in the PIC domain, while the red curve shows the number cell updates in the MHD model. In this particular test, the PIC model uses about 90% of the computational time

7.6.4 MHD-EPIC Applications

MHD-EPIC has been used to simulate the terrestrial magnetosphere [260, 272], the interaction of Mercury [273] and Mars [274] with the solar wind, and the mini-magnetosphere of Ganymede [244, 245]. To demonstrate the capabilities of MHD-EPIC here, we show some of the Earth magnetosphere simulation results [260].

Flux transfer events (FTEs). The iPIC3D code is initialized from a steady Hall MHD state. At the beginning of the coupled simulation, a reconnection X line already exists near the equatorial plane along the dayside magnetopause. The PIC code inherits the magnetic field topology and starts evolving based on Maxwell's equations and the motion of the macroparticles. An overview of the evolution of the dayside magnetopause is shown in Fig. 7.16, which contains the Hall magnetic field B_y and the field lines at the meridional plane inside the PIC box. At $t = 70$ s, B_y has already increased to about 8 nT. The Hall field extends far away from the X line with roughly the same field strength for each branch. Fifteen seconds later, south of the existing reconnection point, another X line starts to form at around $x = 10.2R_E$ and $z = -1R_E$. At $t = 145$ s, both X lines can be seen clearly, and a flux rope-like structure forms between the two X lines. The top X line has moved to about $z = 0.5R_E$. The bottom X line is almost steady to this point, but

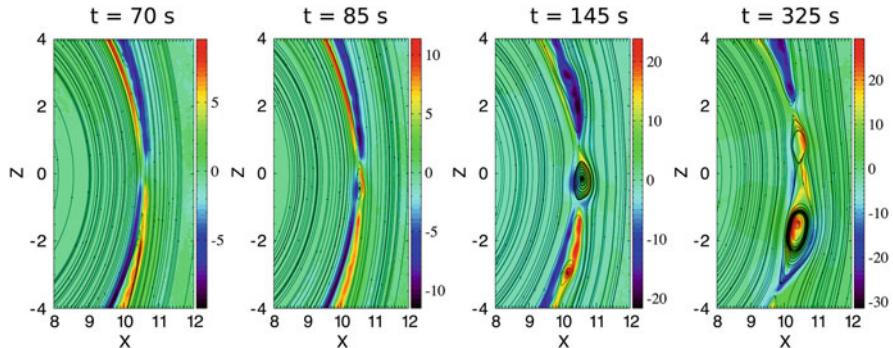


Fig. 7.16 A series of snapshots showing B_y strength (color) and the projected magnetic field lines in the meridional plane inside the PIC region. The color bar is different in each plot (reproduced from [260] with the permission of AGU)

it will move southward later. At $t = 325$ s, the top and bottom X lines reach about $z = 1.8R_E$ and $z = -3.5R_E$, respectively, and the center of the flux rope is moving southward with the bottom X line. Since the flux rope is moving away from the top X line, the current sheet between them becomes unstable, and a secondary flux rope is generated (rightmost panel of Fig. 7.16). During the 1 h simulation, flux ropes form near the subsolar point and move toward the poles quasi-periodically.

More complicated structures arise in 3D, as one see in Fig. 4 of Chen et al. [260]. Along the flux rope, the ion velocity varies. Close to the duskside (positive y), the flux rope moves slowly, because the northward reconnection jet produced by the second X line slows down the southward flow originating from the primary X line. Close to the dawnside (negative y), the flux rope moves faster, because the reconnection at the second X line is not strong enough to significantly slow the southward flow ejected from the primary X line. It was found that (i) flux ropes arise from multiple X line reconnection and can grow in time along the dawn-dusk direction, (ii) the poleward moving velocity varies along a flux rope and makes them tilted, and (iii) two flux ropes can merge and form a new long flux rope.

Crescent distribution. Crescent-shape electron phase space distribution has been observed near the electron diffusion region at the dayside magnetopause by MMS [275]. The same distribution is also found in the 3D MHD-EPIC simulation. The phase space distribution of electrons inside a cube region, $10.27R_E < x < 10.33R_E$, $-0.3R_E < y < 0.3R_E$, and $-2.1R_E < z < -1.9R_E$, is shown in Fig. 7.17. The crescent distribution is found in the $V_y - V_x$ plane, corresponding to the two velocity components perpendicular to the magnetic field. The crescent hot electrons are drifting along negative y direction with a speed close to 3,000 km/s. The direction of the flow is consistent with the $E \times B$ direction, and the velocity of the crescent particles is very close to the MMS observation [275]. Slightly farther away from the reconnection site, where the Larmor field appears, inside a cube of $10.08R_E < x < 10.14R_E$, $-0.3R_E < y < 0.3R_E$, and $-2.1R_E < z < -1.9R_E$,

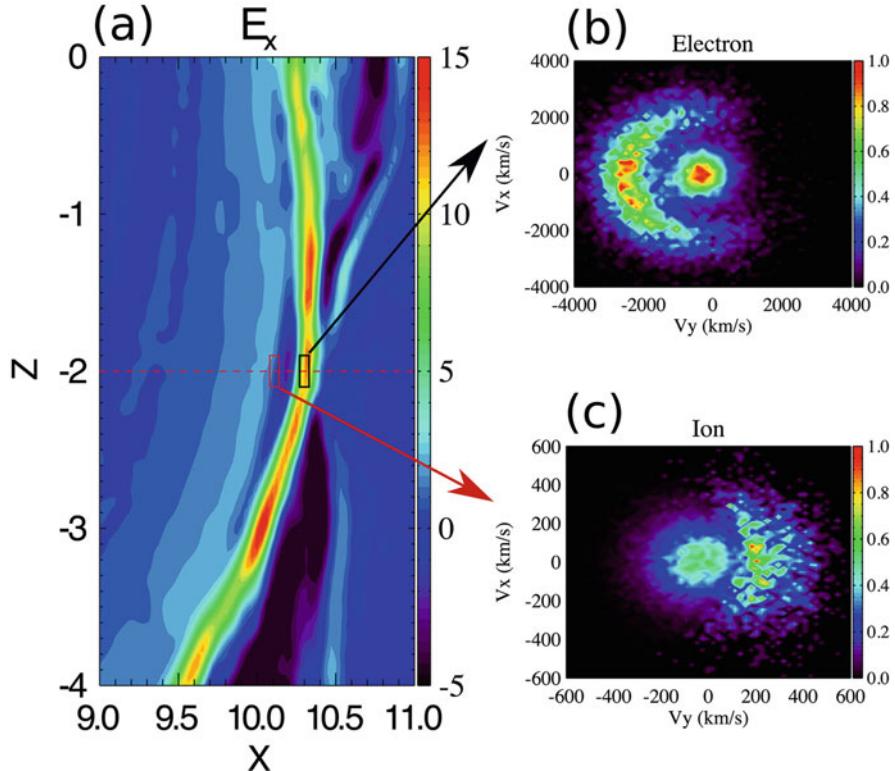


Fig. 7.17 Crescent electron and ion phase space distributions (a) E_x (mV/m) in the meridional plane at $t = 3600$ s, (b) normalized electron distribution in $V_y - V_x$ phase space, and (c) ion phase space distribution. The phase space density is normalized (reproduced from [260] with the permission of AGU)

the ion phase space distribution also presents crescent-like shape as it is shown in Fig. 7.17c. The crescent ions drift in positive y direction because E_x is negative.

Compared to models relying on physical or numerical resistivity to generate FTEs or investigate reconnection process, the MHD-EPIC model makes a significant step forward by incorporating a self-consistent kinetic description of reconnection into a global MHD model. In the kinetic simulation community, there is still debate about the accuracy of the implicit PiC method for magnetic reconnection studies. The results of [269] are quite encouraging, but we need further investigations to prove beyond reasonable doubt that validity of implicit PiC codes for heliophysics applications.

In general, the MHD-EPIC model offers a powerful tool to study magnetospheric physics. A future application is covering the tail reconnection site with another PIC region, so that both dayside and tail reconnections are handled by a kinetic code so that one can study substorms in a more realistic way.

7.7 Summary

Over the last quarter of century, global magnetosphere models came a long way. Some 50 years ago, when the United States placed two astronauts on the moon, all the compute power available to NASA was considerably less than that of a typical smart phone today. Computational methods were also simplistic by today's standards. Even though Sergei Godunov published his seminal paper [276] a decade earlier, it was still several years before Bram van Leer showed how to construct higher-order Godunov schemes [277–281] and more than a decade before the publication of the Roe solver [282]. In the 1980s and 1990s, modern computational methods revolutionized computational plasma physics resulting in the development of the LFM code [6, 108] and BATS-R-US [29]. The performance of supercomputers also underwent qualitative leaps. Compared to first vector supercomputer, today's top machine performs over a million times more operations per second. The combination of hardware, software, and algorithmic advances completely revolutionized space plasma modeling. Simulations that were beyond reach only a decade ago can be routinely performed, and the physical validity of these simulations has greatly improved.

However, nature is so multiscale and so complex that even these huge advances are not enough to carry out full physics simulations of space plasmas. We still need to choose our simplifying assumptions, numerical technologies, grid adaptation, and parallelization wisely to be able to describe space plasma phenomena.

In this paper, we outlined the state of the art of adaptive global MHD simulations of geospace plasmas. We used the SWMF/BATS-R-US and MHD-EPIC codes as examples to demonstrate the capabilities and challenges. While these are not the only modern, high-performance global space plasma simulation codes, they represent cutting-edge technologies in adaptive physics, adaptive grid, and adaptive time-stepping technologies.

Acknowledgments This work was supported by the NSF Strategic Capabilities grant AGS-1322543, by the NASA Heliophysics Strategic Capabilities grant 80NSSC19K0373, and by the NASA Magnetospheric Multi-Scale mission contract 80NSSC19K0564.

References

1. J.N. LeBoeuf, T. Tajima, C.F. Kennel, J.M. Dawson, Geophys. Res. Lett. **8**, 257 (1981). <https://doi.org/10.1029/gl008i003p00257>
2. C.C. Wu, R. Walker, J.M. Dawson, Geophys. Res. Lett. **8**, 523 (1981). <https://doi.org/10.1029/GL008i005p00523>
3. S. Brecht, J. Lyon, J. Fedder, K. Hain, Geophys. Res. Lett. **8**, 397 (1981). <https://doi.org/10.1029/GL008i004p00397>
4. S. Brecht, J. Lyon, J. Fedder, K. Hain, J. Geophys. Res. **87**(A8), 6098 (1982). <https://doi.org/10.1029/ja087ia08p06098>
5. J.G. Lyon, J. Fedder, J. Huba, J. Geophys. Res. **91**(A7), 8057 (1986). <https://doi.org/10.1029/ja091ia07p08057>

6. J.G. Lyon, J.A. Fedder, C.M. Mobarry, J. Atmos. Sol. Terr. Phys. **66**(15–16), 1333 (2004). <https://doi.org/10.1016/j.jastp.2004.03.020>
7. J. Raeder, J. Berchem, M. Ashour-Abdalla, in *The Physics of Space Plasmas*, vol. 14, ed. by T. Chang, J.R. Jasperse (MIT Center For Theoretical Geo/Cosmo Plasma Physics, Cambridge, 1996), p. 403
8. J. Raeder, R.J. Walker, M. Ashour-Abdalla, Geophys. Res. Lett. **22**(4), 349 (1995). <https://doi.org/10.1029/94gl03380>
9. K. Watanabe, T. Sato, J. Geophys. Res. **95**, 75 (1990)
10. A. Usadi, A. Kageyama, K. Watanabe, T. Sato, J. Geophys. Res. **98**, 7503 (1993)
11. P. Janhunen, in *Proceedings of the ESA 1996 Symposium on Environment Modelling for Space-Based Applications*. ESA Special Publication, vol. 392, ed. by T.D. Guyenne, A. Hilgers (Esa Sp-392, 1996), pp. 233–239
12. W.W. White, G.L. Siscoe, G.M. Erickson, Z. Kaymaz, N.C. Maynard, K.D. Siebert, B.U.Ö. Sonnerup, D.R. Weimer, Geophys. Res. Lett. **25**(10), 1605 (1998)
13. G.L. Siscoe, N.U. Crooker, G.M. Erickson, B.U.Ö. Sonnerup, K.D. Siebert, D.R. Weimer, W.W. White, N.C. Maynard, Geophys. Monogr. **118**, 41 (2000). <https://doi.org/10.1029/GM118p0041>
14. J.A. Linker, Z. Mikić, D.D. Schnack, in *Proceedings of the Third SOHO Workshop* (European Space Agency, Estes Park, 1994), pp. 249–252
15. J.A. Linker, Z. Mikić, D.A. Biesecker, R.J. Forsyth, S.E. Gibson, A.J. Lazarus, A. Lecinski, P. Riley, A. Szabo, B.J. Thompson, J. Geophys. Res. **104**, 9809 (1999). <https://doi.org/10.1029/1998ja000159>
16. A.V. Usmanov, Sol. Phys. **146**, 377 (1993). <https://doi.org/10.1007/bf00662021>
17. A.V. Usmanov, M.L. Goldstein, B.P. Besser, J.M. Fritz, J. Geophys. Res. **105**, 12675 (2000). <https://doi.org/10.1029/1999ja000233>
18. K. Hayashi, J. Geophys. Res. **118**, 6889 (2013). <https://doi.org/10.1002/2013ja018991>
19. X. Feng, L. Yang, C. Xiang, S.T. Wu, Y. Zhou, D. Zhong, Astrophys. J. **723**, 300 (2010). <https://doi.org/10.1088/0004-637x/723/1/300>
20. D. Odstrčil, Adv. Space Res. **32**(4), 497 (2003)
21. D. Odstrčil, V.J. Pizzo, Sol. Phys. **259**, 297 (2009). <https://doi.org/10.1007/s11207-009-9449-z>
22. N.V. Pogorelov, G.P. Zank, T. Ogino, Astrophys. J. **614**(2), 1007 (2004). <https://doi.org/10.1086/423798>
23. J.M. Stone, J.F. Hawley, C.R. Evans, M.L. Norman, Astrophys. J. **388**(2), 415 (1992)
24. T. Ogino, J. Geophys. Res. **91**, 6791 (1986). <https://doi.org/10.1029/ja091ia06p06791>
25. H. Washimi, T. Tanaka, Space Sci. Rev. **78**, 85 (1996)
26. R.M. Winglee, Geophys. Res. Lett. **25**, 4441 (1998)
27. R.M. Winglee, W. Lewis, G. Lu, J. Geophys. Res. **110**, A12s24 (2005). <https://doi.org/10.1029/2004ja010909>
28. G. Toth, Astroph. Lett. Comm. **34**, 245 (1996)
29. K. Powell, P. Roe, T. Linde, T. Gombosi, D.L. De Zeeuw, J. Comput. Phys. **154**, 284 (1999). <https://doi.org/10.1006/jcph.1999.6299>
30. G. Toth, B. van der Holst, I.V. Sokolov, D.L. De Zeeuw, T.I. Gombosi, F. Fang, W.B. Manchester, X. Meng, D. Najib, K.G. Powell, Q.F. Stout, A. Glocer, Y.J. Ma, M. Opher, J. Comput. Phys. **231**(3), 870 (2012). <https://doi.org/10.1016/j.jcp.2011.02.006>
31. G. Toth, I.V. Sokolov, T.I. Gombosi, D.R. Chesney, C.R. Clauer, D.L. de Zeeuw, K.C. Hansen, K.J. Kane, W.B. Manchester, R.C. Oehmke, K.G. Powell, A.J. Ridley, I.I. Roussev, Q.F. Stout, O. Volberg, R.A. Wolf, S. Sazykin, A. Chan, B. Yu, J. Kóta, J. Geophys. Res. **110**(A12), A12226 (2005). <https://doi.org/10.1029/2005ja011126>
32. J. Raeder, R.L. MacPherron, L.A. Frank, S. Kokubun, G. Lu, T. Mukai, W.R. Paterson, J.B. Sigwarth, H.J. Singer, J.A. Slavin, J. Geophys. Res., 381 (2001)
33. C.C. Goodrich, A.L. Sussman, J.G. Lyon, M.A. Shay, P.A. Cassak, J. Atmos. Solar Terr. Phys. **66**, 1469 (2004)

34. J.G. Luhmann, S.C. Solomon, J.A. Linker, J.G. Lyon, Z. Mikic, D. Odstrcil, W. Wang, M. Wiltberger, *J. Atmos. Solar Terr. Phys.* **66**, 1243 (2004). <https://doi.org/10.1016/j.jastp.2004.04.005>
35. M. Ashour-Abdalla, G. Lapenta, R.J. Walker, M. El-Alaoui, H. Liang, *J. Geophys. Res.* **120** (2015). <https://doi.org/10.1002/2014ja020316>
36. L.K. Daldorff, G. Toth, T.I. Gombosi, G. Lapenta, J. Amaya, S. Markidis, J.U. Brackbill, *J. Comput. Phys.* **268**, 236 (2014). <https://doi.org/10.1016/j.jcp.2014.03.009>
37. L.M. Noble, F.L. Scarf, *Astrophys. J.* **138**, 1169 (1963). <https://doi.org/10.1086/147715>
38. F.L. Scarf, L.M. Noble, *Astrophys. J.* **141**, 1479 (1965). <https://doi.org/10.1086/148236>
39. Y.C. Whang, C.C. Chang, *J. Geophys. Res.* **70**, 4175 (1965). <https://doi.org/10.1029/JZ070i017p04175>
40. P.A. Sturrock, R.E. Hartle, *Phys. Rev. Lett.* **16**, 628 (1966). <https://doi.org/10.1103/PhysRevLett.16.628>
41. K.H. Schatten, *Nature* **220**, 1211 (1968). <https://doi.org/10.1038/2201211a0>
42. K.H. Schatten, *Nature* **222**, 652 (1969). <https://doi.org/10.1038/222652a0>
43. K.H. Schatten, J.M. Wilcox, N.F. Ness, *Sol. Phys.* **6**, 442 (1969). <https://doi.org/10.1007/bf00146478>
44. M.D. Altschuler, G. Newkirk, *Sol. Phys.* **9**(1), 131 (1969). <https://doi.org/10.1007/bf00145734>
45. G. Newkirk, M.D. Altschuler, *Sol. Phys.* **13**, 131 (1970). <https://doi.org/10.1007/bf00963948>
46. K.H. Schatten, *Cosmic Electrodyn.* **2**, 232 (1971)
47. T.I. Gombosi, B. van der Holst, W.B. Manchester, I.V. Sokolov, *Liv. Rev. Sol. Phys.* **15**(1), 4 (2018). <https://doi.org/10.1007/s41116-018-0014-4>
48. J.T. Hoeksema, J.M. Wilcox, P.H. Scherrer, *J. Geophys. Res.* **87**, 10331 (1982). <https://doi.org/10.1029/JA087iA12p10331>
49. G. Toth, X. Meng, T.I. Gombosi, A. Ridley, *J. Geophys. Res.* **116**, A07211 (2011). <https://doi.org/10.1029/2010ja016370>
50. G.W. Pneuman, *Astrophys. J.* **145**, 242 (1966). <https://doi.org/10.1086/148759>
51. G.W. Pneuman, *Sol. Phys.*, 578–597 (1968). <https://doi.org/10.1007/bf00151939>
52. G.W. Pneuman, *Sol. Phys.* **6**, 255 (1969). <https://doi.org/10.1007/bf00150951>
53. G.W. Pneuman, R.A. Kopp, *Sol. Phys.* **18**, 258 (1971)
54. E.N. Parker, *Interplanetary Dynamical Processes* (Wiley-Interscience, New York, 1963)
55. R.S. Steinolfson, M. Dryer, Y. Nakagawa, *J. Geophys. Res.* **80**, 1223 (1975). <https://doi.org/10.1029/JA080i010p01223>
56. R.S. Steinolfson, S.T. Wu, M. Dryer, E. Tandberg-Hanssen, *Astrophys. J.* **225**, 259 (1978). <https://doi.org/10.1086/156489>
57. V. Pizzo, *J. Geophys. Res.* **83**(A12), 5563 (1978)
58. V.J. Pizzo, *J. Geophys. Res.* **85**(A2), 727 (1980)
59. V.J. Pizzo, *J. Geophys. Res.* **87**(A6), 4374 (1982)
60. R.S. Steinolfson, S.T. Suess, S.T. Wu, *Astrophys. J.* **255**, 730 (1982). <https://doi.org/10.1086/159872>
61. R.S. Steinolfson, *J. Geophys. Res.* **93**, 14261 (1988)
62. R.S. Steinolfson, A.J. Hundhausen, *J. Geophys. Res.* **93**, 14269 (1988)
63. V.J. Pizzo, *J. Geophys. Res.* **94**(A7), 8673 (1989)
64. R.S. Steinolfson, *J. Geophys. Res.* **95**, 20693 (1990)
65. V.J. Pizzo, *J. Geophys. Res.* **96**(A4), 5405 (1991)
66. R.S. Steinolfson, *J. Geophys. Res.* **97**, 10,811 (1992)
67. V.J. Pizzo, K.B. MacGregor, P.B. Kunasz, *Astrophys. J.* **404**, 788 (1993). <https://doi.org/10.1086/172333>
68. V.J. Pizzo, J.T. Gosling, *Geophys. Res. Lett.* **21**, 2063 (1994)
69. V.J. Pizzo, *J. Geophys. Res.* **99**(A3), 4185 (1994)
70. V.J. Pizzo, *J. Geophys. Res.* **99**(A3), 4173 (1994)
71. R.S. Steinolfson, *Space Sci. Rev.* **70**, 289 (1994)
72. Y.M. Wang, N.R. Sheeley, Jr., *Astrophys. J.* **355**, 726 (1990). <https://doi.org/10.1086/168805>

73. L.A. Fisk, N.A. Schwadron, T.H. Zurbuchen, *Space Sci. Rev.* **86**, 51 (1998). <https://doi.org/10.1023/a:1005015527146>
74. R. Lionello, P. Riley, J.A. Linker, Z. Mikić, *Astrophys. J.* **625**(1), 463 (2005). <https://doi.org/10.1086/429268>
75. A.F. Rappazzo, W.H. Matthaeus, D. Ruffolo, S. Servidio, M. Velli, *Astrophys. J. Lett.* **758**(1), L14 (2012). <https://doi.org/10.1088/2041-8205/758/1/L14>
76. V.S. Titov, Z. Mikic, T. Török, J.A. Linker, O. Panasenco, *Astrophys. J.* **759**, 70 (2012). <https://doi.org/10.1088/0004-637x/759/1/70>
77. S.K. Antiochos, J.A. Linker, R. Lionello, Z. Mikić, V. Titov, T.H. Zurbuchen, *Space Sci. Rev.* **172**(1-4), 169 (2012). <https://doi.org/10.1007/s11214-011-9795-7>
78. D. Odstrčil, V.J. Pizzo, *J. Geophys. Res.* **104**, 28,225 (1999)
79. C. Arge, J. Luhmann, D. Odstrčil, C. Schrijver, Y. Li, J. *Atmos. Solar Terr. Phys.* **66**(15), 1295 (2004). <https://doi.org/10.1016/j.jastp.2004.03.018>
80. D. Odstrčil, J.A. Linker, R. Lionello, Z. Mikić, P. Riley, V.J. Pizzo, J.G. Luhmann, *J. Geophys. Res.* **107**, 1493 (2002). <https://doi.org/10.1029/2002ja009334>
81. D. Odstrčil, V.J. Pizzo, J.A. Linker, P. Riley, R. Lionello, Z. Mikic, *J. Atmos. Solar Terr. Phys.* **66**, 1311 (2004)
82. K. Hayashi, *J. Geophys. Res.* **117**, A08105 (2012). <https://doi.org/10.1029/2011ja017490>
83. B.V. Jackson, P.L. Hick, M. Kojima, A. Yokobe, *J. Geophys. Res.* **103**(12), 12049 (1998)
84. V.G. Merkin, J.G. Lyon, S.L. McGregor, D.M. Pahud, *Geophys. Res. Lett.* **38**(14), L14107 (2011). <https://doi.org/10.1029/2011gl047822>
85. V.G. Merkin, J.G. Lyon, D. Lario, C.N. Arge, C.J. Henney, *J. Geophys. Res. Space Phys.* **121**(4), 2866 (2016). <https://doi.org/10.1002/2015ja022200>
86. B. Zhang, K.A. Sorathia, J.G. Lyon, V.G. Merkin, M. Wiltsberger, preprint, arXiv:1810.10861 (2018)
87. D. Shiota, R. Kataoka, Y. Miyoshi, T. Hara, C. Tao, K. Masunaga, Y. Futaana, N. Terada, *Space Weather* **12**(4), 187 (2014). <https://doi.org/10.1002/2013sw000989>
88. D. Shiota, R. Kataoka, *Space Weather* **14**, 56 (2016). <https://doi.org/10.1002/2015sw001308>
89. T.K. Kim, N.V. Pogorelov, G.P. Zank, H.A. Elliott, D.J. McComas, *Astrophys. J.* **832**(1), 72 (2016). <https://doi.org/10.3847/0004-637x/832/1/72>
90. J. Pomoell, S. Poedts, *J. Space Weather Space Clim.* **8**, A35 (2018). <https://doi.org/10.1051/swsc/2018020>
91. S.R. Cranmer, A.A. Van Ballegooijen, *Astrophys. J.* **720**(1), 824 (2010). <https://doi.org/10.1088/0004-637x/720/1/824>
92. I.V. Sokolov, B. van der Holst, R. Oran, C. Downs, I.I. Roussev, M. Jin, W.B. Manchester, R.M. Evans, T.I. Gombosi, *Astrophys. J.* **764**, 23 (2013). <https://doi.org/10.1088/0004-637x/764/1/23>
93. J.V. Hollweg, *J. Geophys. Res.* **91**(A4), 4111 (1986). <https://doi.org/10.1029/JA091iA04p04111>
94. B. van der Holst, I.V. Sokolov, X. Meng, M. Jin, W.B. Manchester, G. Toth, T.I. Gombosi, *Astrophys. J.* **782**(2), 81 (2014). <https://doi.org/10.1088/0004-637x/782/2/81>
95. R. Oran, B. van der Holst, E. Landi, M. Jin, I.V. Sokolov, T.I. Gombosi, *Astrophys. J.* **778**, 176 (2013). <https://doi.org/10.1088/0004-637x/778/2/176>
96. A.V. Usmanov, M.L. Goldstein, *J. Geophys. Res.* **108**(A9), 1354 (2003). <https://doi.org/10.1029/2002ja009777>
97. Y.Q. Hu, X. Li, S.R. Habbal, *J. Geophys. Res.* **108**, 9 (2003). <https://doi.org/10.1029/2003ja009889>
98. T.K. Suzuki, S.i. Inutsuka, *Astrophys. J. Lett.* **632**, L49 (2005). <https://doi.org/10.1086/497536>
99. A. Verdini, M. Velli, W.H. Matthaeus, S. Oughton, P. Dmitruk, *Astrophys. J. Lett.* **708**(2), L116 (2009). <https://doi.org/10.1088/2041-8205/708/2/L116>
100. T. Matsumoto, T.K. Suzuki, *Astrophys. J.* **749**, 8 (2012). <https://doi.org/10.1088/0004-637x/749/1/8>

101. R. Lionello, M. Velli, C. Downs, J.A. Linker, Z. Mikić, *Astrophys. J.* **796**, 111 (2014). <https://doi.org/10.1088/0004-637x/796/2/111>
102. R. Lionello, M. Velli, C. Downs, J.A. Linker, Z. Mikić, A. Verdini, *Astrophys. J.* **784**, 120 (2014). <https://doi.org/10.1088/0004-637x/784/2/120>
103. J.N. LeBoeuf, T. Tajima, C.F. Kennel, J.M. Dawson, *Geophys. Res. Lett.* **5**(7), 609 (1978). <https://doi.org/10.1029/GL005i007p00609>
104. T. Ogino, R.J. Walker, *Geophys. Res. Lett.* **11**, 1018 (1984)
105. T.I. Gombosi, D.N. Baker, A. Balogh, P.J. Erickson, J.D. Huba, L.J. Lanzerotti, *Space Sci. Rev.*, 1–55 (2017). <https://doi.org/10.1007/s11214-017-0357-5>
106. K.H. Hain, *J. Comput. Phys.* **73**(1), 131 (1987). [https://doi.org/10.1016/0021-9991\(87\)90110-0](https://doi.org/10.1016/0021-9991(87)90110-0)
107. C.R. Evans, J.F. Hawley, *Astrophys. J.* **332**, 659 (1988). <https://doi.org/10.1086/166684>
108. J.G. Lyon, S.H. Brecht, J.D. Huba, J.A. Fedder, P.J. Palmadesso, *Phys. Rev. Lett.* **46**, 1038 (1981). <https://doi.org/10.1103/PhysRevLett.46.1038>. <https://link.aps.org/doi/10.1103/PhysRevLett.46.1038>
109. W.W. White, J.A. Schoendorf, K.D. Siebert, N.C. Maynard, D.R. Weimer, G.L. Wilson, B.U.Ö. Sonnerup, G.L. Siscoe, G.M. Erickson, *Geophys. Monogr.* **125**, 229 (2001). <https://doi.org/10.1029/GM125p0229>
110. T. Ogino, R. Walker, M. Ashour-Abdalla, J. Dawson, *J. Geophys. Res.* **91**, 10029 (1986)
111. K. Watanabe, T. Sato, *Geophys. Res. Lett.* **15**, 717 (1988)
112. T. Tanaka, *J. Comput. Phys.* **111**, 381 (1994)
113. T. Tanaka, *J. Geophys. Res.* **100**(A7), 12057 (1995)
114. J. Raeder, J. Berchem, M. Ashour-Abdalla, *J. Geophys. Res.* **103**(A7), 14,787 (1998). <https://doi.org/10.1029/98ja00014>
115. R.S. Steinolfson, R.M. Winglee, *J. Geophys. Res.* **98**, 7537 (1993)
116. P. Janhunen, M. Palmroth, T. Laitinen, I. Honkonen, L. Juusola, G. Fácskó, T. Pulkkinen, *J. Atmos. Solar Terr. Phys.* **80**, 48 (2012). <https://doi.org/10.1016/j.jastp.2012.03.006>
117. K.G. Powell, P.L. Roe, R.S. Myong, T.I. Gombosi, D.L. DeZeeuw, in *12th Computational Fluid Dynamics Conference* (Aiaa-95-1704, San Diego, 1995), p. 661. <https://doi.org/10.2514/6.1995-1704>
118. J. Raeder, *J. Geophys. Res.* **104**(A8), 17357 (1999)
119. T. Gombosi, K. Powell, B. van Leer, *J. Geophys. Res.* **105**(A6), 13141 (2000). <https://doi.org/10.1029/1999ja000342>
120. J. Raeder, *J. Geophys. Res.* **105**, 13,149 (2000)
121. G. Toth, D.L. De Zeeuw, T.I. Gombosi, K.G. Powell, *J. Comput. Phys.* **217**, 722 (2006). <https://doi.org/10.1016/j.jcp.2006.01.029>
122. Y.Q. Hu, X.C. Guo, C. Wang, *J. Geophys. Res.* **112**(A7), A07215 (2007). <https://doi.org/10.1029/2006ja012145>
123. C. Wang, X. Guo, Z. Peng, B. Tang, T. Sun, W. Li, Y. Hu, *Sci. China Earth Sci.* **56**(7), 1141 (2013). <https://doi.org/10.1007/s11430-013-4608-3>
124. CCMC. Community Coordinated Modeling Center. <http://ccmc.gsfc.nasa.gov>
125. V.G. Merkin, J.G. Lyon, *J. Geophys. Res.* **115**(A10), A10202 (2010). <https://doi.org/10.1029/2010ja015461>
126. A.D. Richmond, E.C. Ridley, R.G. Roble, *Geophys. Res. Lett.* **19**(6), 601 (1992). <https://doi.org/10.1029/92gJ00401>
127. M. Wiltberger, W. Wang, A. Burns, S. Solomon, J. Lyon, C. Goodrich, *J. Atmos. Solar Terr. Phys.* **66**, 1411 (2004). <https://doi.org/10.1016/j.jastp.2004.03.026>
128. W. Wang, M. Wiltberger, A. Burns, S. Solomon, T. Killeen, N. Maruyama, J. Lyon, *J. Atmos. Solar Terr. Phys.* **66**(15–16), 1425 (2004). <https://doi.org/10.1016/j.jastp.2004.04.008>
129. A. Harten, *J. Comput. Phys.* **49**(3), 357 (1983). [https://doi.org/10.1016/0021-9991\(83\)90136-5](https://doi.org/10.1016/0021-9991(83)90136-5)
130. P. Janhunen, H.E.J. Koskinen, *Geophys. Res. Lett.* **24**(11), 1419 (1997)
131. M. Liemohn, N.Y. Ganushkina, D.L. De Zeeuw, L. Rastaetter, M. Kuznetsova, D.T. Welling, G. Toth, R. Ilie, T.I. Gombosi, B. van der Holst, *Space Weather* **16**(10), 1583 (2018). <https://doi.org/10.1029/2018sw001953>

132. A.J. Ridley, T.I. Gombosi, D.L. DeZeeuw, Ann. Geophys. **22**, 567 (2004). <https://doi.org/10.5194/angeo-22-567-2004>
133. F. Toffoletto, S. Sastrykin, R. Spiro, R. Wolf, Space Sci. Rev. **107**, 175 (2003). <https://doi.org/10.1023/a:1025532008047>
134. A. Glocer, M. Fok, X. Meng, G. Toth, N. Buzulukova, S. Chen, K. Lin, J. Geophys. Res. **118**(4), 1635 (2013). <https://doi.org/10.1002/jgra.50221>
135. M.C. Fok, N.Y. Buzulukova, S.H. Chen, A. Glocer, T. Nagai, P. Valek, J.D. Perez, J. Geophys. Res. **119**(9), 7522 (2014). <https://doi.org/10.1002/2014ja020239>
136. D.L. De Zeeuw, S. Sastrykin, R.A. Wolf, T.I. Gombosi, A.J. Ridley, Toth, J. Geophys. Res. **109**(A12), A12219 (2004). <https://doi.org/10.1029/2003ja010366>
137. A. Glocer, G. Toth, M. Fok, T. Gombosi, M. Liemohn, J. Atmos. Solar Terr. Phys. **71**(16), 1653 (2009). <https://doi.org/10.1016/j.jastp.2009.01.003>
138. T.I. Gombosi, G. Toth, D.L. De Zeeuw, K.C. Hansen, K. Kabin, K.G. Powell, J. Comput. Phys. **177**, 176 (2002). <https://doi.org/10.1006/jcph.2002.7009>
139. X. Meng, G. Toth, T.I. Gombosi, J. Comput. Phys. **231**, 3610 (2012). <https://doi.org/10.1016/j.jcp.2011.12.042>
140. G. Toth, Y.J. Ma, T.I. Gombosi, J. Comput. Phys. **227**, 6967 (2008). <https://doi.org/10.1016/j.jcp.2008.04.010>
141. Y. Ma, A.F. Nagy, K.C. Hansen, D.L. De Zeeuw, T.I. Gombosi, K. Powell, J. Geophys. Res. **107**(A10), 1282 (2002). <https://doi.org/10.1029/2002ja009293>
142. A. Glocer, G. Tóth, Y. Ma, T. Gombosi, J.C. Zhang, L.M. Kistler, J. Geophys. Res. **114**(A12), A12203 (2009). <https://doi.org/10.1029/2009ja014418>
143. Z. Huang, G. Toth, B. van der Holst, Y. Chen, T.I. Gombosi, J. Comput. Phys. **387**, 134 (2019). <https://doi.org/10.1016/j.jcp.2019.02.023>
144. T. Gombosi, D.L. De Zeeuw, K. Powell, A. Ridley, I. Sokolov, Q. Stout, G. Toth, in *Space Plasma Simulation*, ed. by J. Büchner, C.T. Dum, M. Scholer, No. 615 in Lecture Notes in Physics (Springer, Berlin, 2003), p. 251. https://doi.org/10.1007/3-540-27039-6_36
145. J.M. Burgers, *Flow Equations for Composite Gases* (Academic Press, New York, 1969)
146. R.W. Schunk, A.F. Nagy, Rev. Geophys. Space Phys. **18**, 813 (1980)
147. T. Gombosi, C. Rasmussen, J. Geophys. Res. **96**, 7759 (1991). <https://doi.org/10.1029/91ja00012>
148. T.I. Gombosi, *Physics of the Space Environment* (Cambridge University Press, Cambridge, 1998). <https://doi.org/10.1017/cbo9780511529474>
149. U. Shumlak, J. Loverich, J. Comput. Phys. **187**(2), 620 (2003). [https://doi.org/10.1016/s0021-9911\(03\)00151-7](https://doi.org/10.1016/s0021-9911(03)00151-7)
150. G.F. Chew, M.L. Goldberger, F.E. Low, Proc. R. Soc. Lond. A Math. Phys. Eng. Sci. **236**(1204), 112 (1956). <https://doi.org/10.1098/rspa.1956.0116>
151. I.I. Roussev, T.I. Gombosi, I.V. Sokolov, M. Velli, W. Manchester, D.L. DeZeeuw, P. Liewer, G. Toth, J.G. Luhmann, Astrophys. J. **595**, L57 (2003)
152. C. Downs, I.I. Roussev, B. van der Holst, N. Lugaz, I.V. Sokolov, T.I. Gombosi, Astrophys. J. **712**(2), 1219 (2010). <https://doi.org/10.1088/0004-637x/712/2/1219>
153. B. van der Holst, W.B. Manchester, R.A. Frazin, A.M. Vásquez, G. Toth, T.I. Gombosi, Astrophys. J. **725**(1), 1373 (2010). <https://doi.org/10.1088/0004-637x/725/1/1373>
154. C. Downs, I.I. Roussev, B. van der Holst, N. Lugaz, I.V. Sokolov, T.I. Gombosi, Astrophys. J. **728**(1) (2011). <https://doi.org/10.1088/0004-637x/728/1/2>
155. M. Jin, W.B. Manchester, B. van der Holst, J.R. Gruesbeck, R.A. Frazin, E. Landi, A.M. Vasquez, P.L. Lamy, A. Llebaria, A. Fedorov, G. Toth, T. Gombosi, Astrophys. J. **774**5, 6 (2012). <https://doi.org/10.1088/0004-637x/745/1/6>
156. R. Oran, E. Landi, B. van der Holst, S.T. Lepri, A.M. Vásquez, F.A. Nuevo, R. Frazin, W. Manchester, I. Sokolov, T.I. Gombosi, Astrophys. J. **806**(1), 55 (2015). <https://doi.org/10.1088/0004-637x/806/1/55>
157. X. Meng, B. van der Holst, G. Toth, T.I. Gombosi, Mon. Not. R. Astron. Soc. **454**(4), 3697 (2015). <https://doi.org/10.1093/mnras/stv2249>

158. J. Szente, G. Toth, W.B.M. IV, B. van der Holst, E. Landi, T.I. Gombosi, C.R. DeVore, S.K. Antiochos, *Astrophys. J.* **834**(2), 123 (2017). <https://doi.org/10.3847/1538-4357/834/2/123>
159. J. Szente, E. Landi, W.B. Manchester, G. Toth, B. van der Holst, T.I. Gombosi, *Astrophys. J. Suppl.* **242**(1), 1 (2019). <https://doi.org/10.3847/1538-4365/ab16d0>
160. B. van der Holst, W.B. Manchester, K.G. Klein, J.C. Kasper, *Astrophys. J.* **872**(2), L18 (2019). <https://doi.org/10.3847/2041-8213/ab04a5>
161. C.P.T. Groth, D.L. De Zeeuw, T.I. Gombosi, K.G. Powell, *Space Sci. Rev.* **87**, 193 (1999)
162. P.L. Israelevich, A.I. Ershkovich, T.I. Gombosi, *Astron. Astrophys.* **362**, 379 (2000)
163. O. Cohen, I.V. Sokolov, I.I. Roussev, C.N. Arge, W.B. Manchester, T.I. Gombosi, R.A. Frazin, H. Park, M.D. Butala, F. Kamalabadi, M. Velli, *Astrophys. J. Lett.* **654**(2), L163 (2006). <https://doi.org/10.1086/511154>
164. O. Cohen, I.V. Sokolov, I.I. Roussev, T.I. Gombosi, *J. Geophys. Res.* **113**(A3), A03104 (2008). <https://doi.org/10.1029/2007ja012797>
165. R.M. Evans, M. Opher, R. Oran, B. van der Holst, I.V. Sokolov, R. Frazin, T.I. Gombosi, A. Vasquez, *Astrophys. J.* **756**(2) (2012). <https://doi.org/10.1088/0004-637x/756/2/155>
166. I.V. Sokolov, B. van der Holst, W.B. Manchester, D.C.S. Ozturk, J. Szente, A.R. Taktakishvili, G. Toth, M. Jin, T.I. Gombosi, Threaded-field-lines model for the low solar corona powered by the Alfvén wave turbulence (2016)
167. C.P.T. Groth, D.L. De Zeeuw, T.I. Gombosi, K.G. Powell, *Adv. Space Res.* **26**(5), 793 (2000). [https://doi.org/10.1016/s0273-1177\(00\)00008-9](https://doi.org/10.1016/s0273-1177(00)00008-9)
168. I.I. Roussev, T.G. Forbes, T.I. Gombosi, I.V. Sokolov, D.L. DeZeeuw, J. Birn, *Astrophys. J. Lett.* **588**, L45 (2003). <https://doi.org/10.1086/375442>
169. W.B. Manchester, T.I. Gombosi, I.I. Roussev, D.L.D. Zeeuw, I.V. Sokolov, K.G. Powell, G. Toth, M. Opher, *J. Geophys. Res.* **109**(A1), 1102 (2004). <https://doi.org/10.1029/2002ja009672>
170. I.I. Roussev, I.V. Sokolov, T.G. Forbes, T.I. Gombosi, M.A. Lee, J.I. Sakai, *Astrophys. J. Lett.* **605**, L73 (2004). <https://doi.org/10.1086/392504>
171. N. Lugaz, W.B. Manchester, T.I. Gombosi, *Astrophys. J.* **634**, 651 (2005). <https://doi.org/10.1086/491782>
172. N. Lugaz, W.B. Manchester, IV, I.I. Roussev, G. Toth, T.I. Gombosi, *Astrophys. J.* **659**(1), 788 (2007). <https://doi.org/10.1086/512005>
173. N. Lugaz, W.B. Manchester, I.I. Roussev, T.I. Gombosi, *J. Atmos. Solar Terr. Phys.* **70**(2-4), 598 (2008). <https://doi.org/10.1016/j.jastp.2007.08.033>
174. W.B. Manchester, A. Vourlidas, G. Toth, N. Lugaz, I.I. Roussev, I.V. Sokolov, T.I. Gombosi, D.L.D. Zeeuw, M. Opher, *Astrophys. J.* **684**(2), 1448–1460 (2008). <https://doi.org/10.1086/590231>
175. B. van der Holst, W. Manchester, IV, I.V. Sokolov, G. Toth, T.I. Gombosi, D. DeZeeuw, O. Cohen, *Astrophys. J.* **693**(2), 1178 (2009). <https://doi.org/10.1088/0004-637x/693/2/1178>
176. O. Cohen, J.J. Drake, V.L. Kashyap, I.V. Sokolov, T.I. Gombosi, *Astrophys. J. Lett.* **723**(1), L64 (2010). <https://doi.org/10.1088/2041-8205/723/1/l64>
177. N. Lugaz, C. Downs, K. Shibata, I.I. Roussev, A. Asai, T. Gombosi, *Astrophys. J.* **738**(2), 127 (2011). <https://doi.org/10.1088/0004-637x/738/2/127>
178. W.B. Manchester, B. van der Holst, G. Toth, T.I. Gombosi, *Astrophys. J.* **756**(1), 81 (2012). <https://doi.org/10.1088/0004-637x/756/1/81>
179. M. Jin, W. Manchester, B. van der Holst, R. Oran, I. Sokolov, G. Toth, T. Gombosi, Y. Liu, X. Sun, *Astrophys. J.* **773**(1), 50 (2013). <https://doi.org/10.1088/0004-637x/773/1/50>
180. M. Jin, W.B. Manchester, B. van der Holst, I. Sokolov, G. Toth, A. Vourlidas, C.A. de Koning, T.I. Gombosi, *Astrophys. J.* **834**(2), 172 (2017). <https://doi.org/10.3847/1538-4357/834/2/172>
181. I.V. Sokolov, I.I. Roussev, T.I. Gombosi, M.A. Lee, J. Kóta, T.G. Forbes, W.B. Manchester IV, J.I. Sakai, *Astrophys. J. Lett.* **616**, L171 (2004). <https://doi.org/10.1086/426812>
182. J. Kóta, W.B. Manchester, J.R. Jokipii, D.L. de Zeeuw, and T.I. Gombosi, Simulation of SEP Acceleration and Transport at CME-driven Shocks, *AIP Conference Proceedings* **781**, 201–206 (2005). <https://doi.org/10.1063/1.2032697>
183. W. Manchester, T.I. Gombosi, D.L. De Zeeuw, I.V. Sokolov, I.I. Roussev, K.G. Powell, J. Kóta, G. Toth, T.H. Zurbuchen, *Astrophys. J.* **622**, 1225 (2005). <https://doi.org/10.1086/427768>

184. I.V. Sokolov, I.I. Roussev, L.A. Fisk, M.A. Lee, T.I. Gombosi, J.I. Sakai, *Astrophys. J.* **642**(1), L81 (2006). <https://doi.org/10.1086/504406>
185. I.V. Sokolov, I.I. Roussev, M. Skender, T.I. Gombosi, A.V. Usmanov, *Astrophys. J.* **696**(1), 261 (2009). <https://doi.org/10.1088/0004-637x/696/1/261>
186. D. Borovikov, I.V. Sokolov, I.I. Roussev, A. Taktakishvili, T.I. Gombosi, *Astrophys. J.* **864**(1), 88 (2018). <https://doi.org/10.3847/1538-4357/aad68d>
187. T.J. Linde, T.I. Gombosi, P.L. Roe, K.G. Powell, D.L. De Zeeuw, *J. Geophys. Res.* **103**(A2), 1889 (1998). <https://doi.org/10.1029/97ja02144>
188. P.L. Israelevich, T.I. Gombosi, A.I. Ershkovich, K.C. Hansen, C. Groth, D.L. De Zeeuw, K.G. Powell, *Astron. Astrophys.* **376**(1), 288 (2001)
189. M. Opher, P. Liewer, T. Gombosi, W.B. Manchester IV, D. De Zeeuw, I. Sokolov, G. Toth, *Astrophys. J.* **591**(1), L61 (2003). <https://doi.org/10.1086/376960>
190. M. Opher, P. Liewer, M. Velli, L. Bettarini, T. Gombosi, W. Manchester IV, D. De Zeeuw, G. Toth, I. Sokolov, *Astrophys. J.* **611**(1), 575 (2004). <https://doi.org/10.1086/422165>
191. M. Opher, E.C. Stone, T.I. Gombosi, *Science* **316**(5826), 875 (2007). <https://doi.org/10.1126/science.1139480>
192. M. Opher, J. Drake, M. Swisdak, B. Zieger, G. Toth, *Astrophys. J. Lett.* **839**(1), L12 (2017). <https://doi.org/10.3847/2041-8213/aa692f>
193. C.P.T. Groth, D.L. De Zeeuw, T.I. Gombosi, K.G. Powell, *J. Geophys. Res.* **105**(A11), 25053 (2000). <https://doi.org/10.1029/2000ja900093>
194. T.I. Gombosi, D.L. De Zeeuw, C.P.T. Groth, K.G. Powell, Q.F. Stout, *J. Atmos. Solar Terr. Phys.* **62**, 1515 (2000). [https://doi.org/10.1016/s1364-6826\(00\)00091-2](https://doi.org/10.1016/s1364-6826(00)00091-2)
195. W.B. Manchester, T.I. Gombosi, I. Roussev, A. Ridley, D.L.D. Zeeuw, I.V. Sokolov, K.G. Powell, G. Toth, *J. Geophys. Res.* **109**(A2), A02107 (2004). <https://doi.org/10.1029/2003ja010150>
196. G. Toth, D.L. De Zeeuw, T.I. Gombosi, W.B. Manchester, A.J. Ridley, I.V. Sokolov, I.I. Roussev, *Space Weather* **5**(6), S06003 (2007). <https://doi.org/10.1029/2006sw000272>
197. M. Jin, W.B. Manchester, B. van der Holst, I. Sokolov, G. Toth, R.E. Mullinix, A. Taktakishvili, A. Chulaki, T.I. Gombosi, *Astrophys. J.* **834**(2), 173 (2017). <https://doi.org/10.3847/1538-4357/834/2/173>. <https://arxiv.org/abs/1605.05360>
198. T.I. Gombosi, D.L. De Zeeuw, C.P.T. Groth, K.G. Powell, P. Song, in *Physics of Space Plasmas*, vol. 15, ed. by T. Chang, J.R. Jasperse (MIT Press, Cambridge, 1998), pp. 121–128
199. P. Song, D.L. De Zeeuw, T.I. Gombosi, C.P.T. Groth, K.G. Powell, *J. Geophys. Res.* **104**(A12), 28361 (1999)
200. T.I. Gombosi, D.L. De Zeeuw, C.P.T. Groth, K.G. Powell, *Adv. Space Res.* **26**(1), 139 (2000). [https://doi.org/10.1016/s0273-1177\(99\)01040-6](https://doi.org/10.1016/s0273-1177(99)01040-6)
201. A. Ridley, *J. Atmos. Solar Terr. Phys.* **62**, 757 (2000)
202. M.M. Kuznetsova, M. Hesse, L. Rastaetter, A. Taktakishvili, G. Toth, D.L. De Zeeuw, A. Ridley, T.I. Gombosi, *J. Geophys. Res.* **112**(A10), A10210 (2007). <https://doi.org/10.1029/2007ja012316>
203. X. Meng, G. Toth, M.W. Liemohn, T.I. Gombosi, A. Runov, *J. Geophys. Res.* **117**, A08216 (2012). <https://doi.org/10.1029/2012ja017791>
204. X. Meng, G. Toth, A. Glocer, M.C. Fok, T.I. Gombosi, *J. Geophys. Res.* **118**, 5639 (2013). <https://doi.org/10.1002/jgra.50539>
205. D.L. De Zeeuw, T.I. Gombosi, C.P.T. Groth, K.G. Powell, Q.F. Stout, *IEEE Trans. Plasma Sci.* **28**, 1956 (2000). <https://doi.org/10.1109/27.902224>
206. G. Toth, X. Meng, T.I. Gombosi, L. Rastatter, *J. Geophys. Res.* **119**, 310–321 (2014). <https://doi.org/10.1002/2013ja019456>
207. M.W. Liemohn, D.L. De Zeeuw, R. Ilie, N.Y. Ganushkina, *Geophys. Res. Lett.* **38**, L20106 (2011). <https://doi.org/10.1029/2011gl049611>
208. J.D. Haiducek, D.T. Welling, N.Y. Ganushkina, S.K. Morley, D.S. Ozturk, *Space Weather* **15**(12), 1567 (2017). <https://doi.org/10.1002/2017sw001695>

209. K. Kabin, T.I. Gombosi, D.L. De Zeeuw, K.G. Powell, *Icarus* **143**, 397 (2000). <https://doi.org/10.1006/icar.1999.6252>
210. K. Kabin, M.H. Heimpel, R. Rankin, J.M. Aurnou, N. Gomez-Perez, J. Paral, T.I. Gombosi, T.H. Zurbuchen, P.L. Koehn, D.L. DeZeeuw, *Icarus* **195**(1), 1 (2008). <https://doi.org/10.1016/j.icarus.2007.11.028>
211. X. Jia, J.A. Slavin, T.I. Gombosi, L.K.S. Daldorff, G. Toth, B. van der Holst, *J. Geophys. Res.* **120**(6), 4763 (2015). <https://doi.org/10.1002/2015ja021143>
212. X. Jia, J.A. Slavin, G. Poh, G.A. DiBraccio, G. Toth, Y. Chen, J.M. Raines, T.I. Gombosi, *J. Geophys. Res.* **124**(1), 229 (2019). <https://doi.org/10.1029/2018ja026166>
213. R. Bauske, A.F. Nagy, T.I. Gombosi, D.L. De Zeeuw, K.G. Powell, J.G. Luhmann, *J. Geophys. Res.* **103**(A10), 23625 (1998). <https://doi.org/10.1029/98ja01791>
214. Y. Ma, A.F. Nagy, C. Russell, R.J. Strangeway, H.Y. Wei, G. Toth, *J. Geophys. Res.* **118**, 321 (2013). <https://doi.org/10.1029/2012ja018265>
215. Y. Liu, A.F. Nagy, C.P.T. Groth, D.L. De Zeeuw, T.I. Gombosi, K.G. Powell, *Geophys. Res. Lett.* **26**(17), 2689 (1999). <https://doi.org/10.1029/1999gl1900584>
216. R. Bauske, A.F. Nagy, D.L. Dezeeuw, T.I. Gombosi, K.G. Powell, *Adv. Space Res.* **26**, 1571 (2000). [https://doi.org/10.1016/s0273-1177\(00\)00105-8](https://doi.org/10.1016/s0273-1177(00)00105-8)
217. M.W. Liemohn, *J. Geophys. Res.* **111**, A11s01 (2006). <https://doi.org/10.1029/2006ja011970>
218. Y. Ma, X. Fang, J.S. Halekas, S. Xu, C.T. Russell, J.G. Luhmann, A.F. Nagy, G. Toth, C.O. Lee, C. Dong, J.R. Espley, J.P. McFadden, D.L. Mitchell, B.M. Jakosky, *Geophys. Res. Lett.* **45**(15), 7248 (2018). <https://doi.org/10.1029/2018gl077707>
219. L.H. Regoli, C. Dong, Y. Ma, E. Dubinin, W.B. Manchester, S.W. Bouger, D.T. Welling, *J. Geophys. Res.* **123**(9), 7370 (2018). <https://doi.org/10.1029/2017ja025117>
220. T. Cravens, J. Waite, T. Gombosi, N. Lugaz, G. Gladstone, B. Mauk, R. MacDowall, *J. Geophys. Res.* **108**, 1465 (2003). <https://doi.org/10.1029/2003ja010050>
221. Y. Sarkango, X. Jia, G. Toth, *J. Geophys. Res.* **124** (2019). <https://doi.org/10.1029/2019ja026787>
222. K.C. Hansen, T.I. Gombosi, D.L. DeZeeuw, C.P.T. Groth, K.G. Powell, *Adv. Space Res.* **26**(10), 1681 (2000). [https://doi.org/10.1016/s0273-1177\(00\)00078-8](https://doi.org/10.1016/s0273-1177(00)00078-8)
223. T.I. Gombosi, K.C. Hansen, *Science* **307**, 1224 (2005). <https://doi.org/10.1126/science.1108226>
224. K.C. Hansen, A.J. Ridley, T.I. Gombosi, G. Hospodarsky, *Geophys. Res. Lett.* **32**(20), L20s06 (2005). <https://doi.org/10.1029/2005gl022835>
225. A. Glocer, T.I. Gombosi, G. Toth, K.C. Hansen, A.J. Ridley, A. Nagy, *J. Geophys. Res.* **112**, A01304 (2007). <https://doi.org/10.1029/2006ja011755>
226. T.I. Gombosi, A.P. Ingersoll, *Science* **327**(5972), 1476 (2010). <https://doi.org/10.1126/science.1179119>
227. B. Zieger, K.C. Hansen, T.I. Gombosi, D.L. De Zeeuw, *J. Geophys. Res.* **115** (2010). <https://doi.org/10.1029/2009ja014951>
228. X. Jia, M.G. Kivelson, T.I. Gombosi, *J. Geophys. Res.* **117**, A04215 (2012). <https://doi.org/10.1029/2011ja017367>
229. X. Jia, K.C. Hansen, T.I. Gombosi, M.G. Kivelson, G. Toth, D.L. DeZeeuw, A.J. Ridley, *J. Geophys. Res.* **117**, A05225 (2012). <https://doi.org/10.1029/2012ja017575>
230. G. Toth, D. Kovács, K.C. Hansen, T.I. Gombosi, *J. Geophys. Res.* **109**(A11), A11210 (2004). <https://doi.org/10.1029/2004ja010406>
231. T.I. Gombosi, D.L. De Zeeuw, R.M. Häberli, K.G. Powell, *J. Geophys. Res.* **101**, 15233 (1996). <https://doi.org/10.1029/96ja01075>
232. R.M. Häberli, T.I. Gombosi, D.L. De Zeuuw, M.R. Combi, K.G. Powell, *Science* **276**, 939 (1997). <https://doi.org/10.1126/science.276.5314.939>
233. T.I. Gombosi, K.C. Hansen, D.L. De Zeeuw, M.R. Combi, K.G. Powell, *Earth Moon Planets* **79**(1/3), 179 (1999). <https://doi.org/10.1023/a:1006289418660>
234. K.C. Hansen, T. Bagdonat, U. Motschmann, C. Alexander, M.R. Combi, T.E. Cravens, T.I. Gombosi, Y. Jia, I.P. Robertson, *Space Sci. Rev.* **128**, 133 (2007). <https://doi.org/10.1007/s11214-006-9142-6>

235. M. Rubin, K.C. Hansen, M.R. Combi, L.K.S. Daldorff, T.I. Gombosi, V.M. Tenishev, *J. Geophys. Res.* **117**, A06227 (2012). <https://doi.org/10.1029/2011ja017300>
236. T.I. Gombosi, *Geophys. Monogr.* **207**, 169 (2015). <https://doi.org/10.1002/9781118842324.ch10>
237. Z. Huang, G. Toth, T.I. Gombosi, X. Jia, M. Rubin, N. Fougere, V. Tenishev, M.R. Combi, A. Bieler, K.C. Hansen, Y. Shou, K. Altwegg, *J. Geophys. Res.* **121**(5), 4247 (2016). <https://doi.org/10.1002/2015ja022333>
238. Z. Huang, G. Toth, T. Gombosi, A. Bieler, M. Combi, K. Hansen, X. Jia, N. Fougere, Y. Shou, T. Cravens, et al., *Mon. Not. R. Astron. Soc.* **462**(Suppl_1), S468 (2016). <https://doi.org/10.1093/mnras/stw3118>
239. M.R. Combi, K. Kabin, T. Gombosi, D. De Zeeuw, K. Powell, *J. Geophys. Res.* **103**(A5), 9071 (1998). <https://doi.org/10.1029/98ja00073>
240. K. Kabin, M.R. Combi, T.I. Gombosi, D.L. De Zeeuw, K.C. Hansen, K.G. Powell, *Planet. Space Sci.* **49**, 337 (2001). [https://doi.org/10.1016/s0032-0633\(00\)00155-0](https://doi.org/10.1016/s0032-0633(00)00155-0)
241. K. Kabin, M.R. Combi, T.I. Gombosi, A.F. Nagy, D.L. De Zeeuw, K.G. Powell, *J. Geophys. Res.* **104**(A9), 19983 (1999). <https://doi.org/10.1029/1999ja900263>
242. Y. Liu, A. Nagy, K. Kabin, M. Combi, D. DeZeeuw, T. Gombosi, K. Powell, *Geophys. Res. Lett.* **27**(12), 1791 (2000). <https://doi.org/10.1029/1999gl003734>
243. M. Rubin, X. Jia, K. Altwegg, M.R. Combi, L.K.S. Daldorff, T.I. Gombosi, K. Khurana, M.G. Kivelson, V.M. Tenishev, G. Toth, B. van der Holst, P. Wurz, *J. Geophys. Res.* **120**, 3503–3524 (2015). <https://doi.org/10.1002/2015ja021149>
244. G. Toth, X. Jia, S. Markidis, B. Peng, Y. Chen, L. Daldorff, V. Tenishev, D. Borovikov, J. Haiducek, T. Gombosi, A. Glocer, J. Dorelli, *J. Geophys. Res.* **121**(2), 1273 (2016). <https://doi.org/10.1002/2015ja021997>
245. H. Zhou, G. Tóth, X. Jia, Y. Chen, S. Markidis, *J. Geophys. Res.* **124** (2019). <https://doi.org/10.1029/2019ja026643>
246. K. Kabin, T.I. Gombosi, D.L. De Zeeuw, K.G. Powell, P.L. Israelevich, *J. Geophys. Res.* **104**(A2), 2451 (1999). <https://doi.org/10.1029/1998ja900080>
247. K. Kabin, P.L. Israelevich, A.I. Ershkovich, F.M. Neubauer, T.I. Gombosi, D.L. De Zeeuw, K.G. Powell, *J. Geophys. Res.* **105**, 10761 (2000). <https://doi.org/10.1029/2000ja900012>
248. A.F. Nagy, Y. Liu, K.C. Hansen, K. Kabin, T. Gombosi, M.. Combi, D.L. De Zeeuw, K.G. Powell, A. Kliore, *J. Geophys. Res.* **106**, 6151 (2001). <https://doi.org/10.1029/2000ja000183>
249. Y.J. Ma, A.F. Nagy, G. Toth, T.E. Cravens, C.T. Russell, T.I. Gombosi, J.E. Wahlund, F.J. Crary, A.J. Coates, C.L. Bertucci, F.M. Neubauer, *Geophys. Res. Lett.* **34**(24), L24s10 (2007). <https://doi.org/10.1029/2007gl031627>
250. Y.D. Jia, C.T. Russell, K.K. Khurana, G. Toth, J.S. Leisner, T.I. Gombosi, *J. Geophys. Res.* **115**, A04214 (2010). <https://doi.org/10.1029/2009ja014630>
251. Y.D. Jia, C.T. Russell, K.K. Khurana, Y.J. Ma, D. Najib, T.I. Gombosi, *J. Geophys. Res.* **115**, A04215 (2010). <https://doi.org/10.1029/2009ja014873>
252. Y.D. Jia, C.T. Russell, K.K. Khurana, Y.J. Ma, W. Kurth, T.I. Gombosi, *J. Geophys. Res.* **115**, A12243 (2010). <https://doi.org/10.1029/2010ja015534>
253. R.P. Drake, F.W. Doss, R.G. McClaren, M.L. Adams, N. Amato, D. Bingham, C.C. Chou, C. DiStefano, K. Fidkowski, B. Fryxell, T.I. Gombosi, M.J. Grosskopf, J.P. Holloway, B. van der Holst, C.M. Huntington, S. Karni, C.M. Krauland, C.C. Kuranz, E. Larsen, B. van Leer, B. Mallick, D. Marion, W. Martin, J.E. Morel, E.S. Myra, V. Nair, K.G. Powell, L. Rauchwerger, P. Roe, E. Rutter, I.V. Sokolov, Q. Stout, B.R. Torralva, G. Toth, K. Thornton, A.J. Visco, *High Energy Density Phys.* **7**(3), 130 (2011). <https://doi.org/10.1016/j.hedp.2011.03.005>
254. B. van der Holst, G. Toth, I.V. Sokolov, K.G. Powell, J.P. Holloway, E.S. Myra, Q. Stout, M.L. Adams, J.E. Morel, S. Karni, B. Fryxell, R.P. Drake, *Astrophys. J. Suppl.* **194**(2), 23 (2011). <https://doi.org/10.1088/0067-0049/194/2/23>
255. B. van der Holst, G. Toth, I.V. Sokolov, L.K.S. Daldorff, K. Powell, R.P. Drake, *High Energy Density Phys.* **8**(2), 161 (2012). <https://doi.org/10.1016/j.hedp.2012.02.001>

256. A. Pulkkinen, L. Rastätter, M. Kuznetsova, H. Singer, C. Balch, D. Weimer, G. Toth, A. Ridley, T. Gombosi, M. Wiltberger, J. Raeder, R. Weigel, Space Weather **11**(6), 369 (2013). <https://doi.org/10.1002/swe.20056>
257. A. Glocer, L. Rastätter, M. Kuznetsova, A. Pulkkinen, H.J. Singer, C. Balch, D. Weimer, D. Welling, M. Wiltberger, J. Raeder, R.S. Weigel, J. McCollough, S. Wing, Space Weather **14**(7), 469 (2016). <https://doi.org/10.1002/2016sw001387>
258. D.T. Welling, A.J. Ridley, Space Weather **8**, S03002 (2010). <https://doi.org/10.1029/2009sw000494>
259. S. Markidis, G. Lapenta, Rizwan-uddin, Math. Comput. Simul. **80**(7), 1509 (2010). <https://doi.org/10.1016/j.matcom.2009.08.038>. <http://www.sciencedirect.com/science/article/pii/S0378475409002444>. Multiscale modeling of moving interfaces in materials
260. Y. Chen, G. Toth, P. Cassak, X. Jia, T.I. Gombosi, J.A. Slavin, S. Markidis, I.B. Peng, V.K. Jordanova, M.G. Henderson, J. Geophys. Res. **122**(10), 10,318 (2017). <https://doi.org/10.1002/2017ja024186>
261. G. Toth, Y. Chen, T.I. Gombosi, P. Cassak, S. Markidis, I.B. Peng, J. Geophys. Res. **122**(10), 10,336 (2017). <https://doi.org/10.1002/2017ja024189>
262. G. Lapenta, J. Comput. Phys. **334**, 349 (2017). <https://doi.org/10.1016/j.jcp.2017.01.002>
263. Y. Chen, G. Toth, J. Comput. Phys. **386**, 632 (2019). <https://doi.org/10.1016/j.jcp.2019.02.032>
264. J.U. Brackbill, D. Forslund, J. Comput. Phys. **46**, 271 (1982). [https://doi.org/10.1016/0021-9991\(82\)90016-x](https://doi.org/10.1016/0021-9991(82)90016-x)
265. J.U. Brackbill, D.W. Forslund, in *Multiple Time Scales*, ed. by J.U. Brackbill, B.I. Cohen (Academic Press, New York, 1986), pp. 271–310
266. J.U. Brackbill, G. Lapenta, Comm. Comput. Phys. **4**, 433 (2008)
267. G. Lapenta, J.U. Brackbill, P. Ricci, Phys. Plasmas **13**, 055904 (2006). <https://doi.org/10.1063/1.2173623>
268. G. Lapenta, in *Space Plasma Simulation: Approaches, Codes and Applications*, ed. by J. Büchner (Springer, Berlin, 2021)
269. P. Ricci, G. Lapenta, J.U. Brackbill, Geophys. Res. Lett. **29**(23), 3 (2002). <https://doi.org/10.1029/2002gl015314>
270. M. Ashour-Abdalla, G. Lapenta, R. Walker, M. El-Alaoui, H. Liang, M. Zhou, J. Berchem, M.L. Goldstein, Geophys. Res. Lett. **43**(12), 6005 (2016). <https://doi.org/10.1002/2016gl069355>
271. R.J. Walker, G. Lapenta, H. Liang, J. Berchem, M. El-Alaoui, M.L. Goldstein, J. Geophys. Res. **123**(10), 8241 (2018). <https://doi.org/10.1029/2018ja025509>
272. V. Jordanova, G. Delzanno, M. Henderson, H. Godinez, C. Jeffery, E. Lawrence, S. Morley, J. Moulton, L. Vernon, J. Woodroffe, T. Brito, M. Engel, C. Meierbachtol, D. Svyatsky, Y. Yu, G. Toth, D. Welling, Y. Chen, J. Haiducek, S. Markidis, J. Albert, J. Birn, M. Denton, R. Horne, J. Atmos. Solar Terr. Phys. **177**, 148 (2018). <https://doi.org/10.1016/j.jastp.2017.11.006>
273. Y. Chen, G. Tóth, X. Jia, J.A. Slavin, W. Sun, S. Markidis, T.I. Gombosi, J.M. Raines, J. Geophys. Res. **124**(11), 8954 (2019). <https://doi.org/10.1029/2019ja026840>
274. Y. Ma, C.T. Russell, G. Toth, Y. Chen, A.F. Nagy, Y. Harada, J. McFadden, J.S. Halekas, R. Lillis, J.E.P. Connerney, J. Espley, G.A. DiBraccio, S. Markidis, I.B. Peng, X. Fang, B.M. Jakosky, J. Geophys. Res. **123**(5), 3742 (2018). <https://doi.org/10.1029/2017ja024729>
275. J.L. Burch, R.B. Torbert, T.D. Phan, L.J. Chen, T.E. Moore, R.E. Ergun, J.P. Eastwood, D.J. Gershman, P.A. Cassak, M.R. Argall, S. Wang, M. Hesse, C.J. Pollock, B.L. Giles, R. Nakamura, B.H. Mauk, S.A. Fuselier, C.T. Russell, R.J. Strangeway, J.F. Drake, M.A. Shay, Y.V. Khotyaintsev, P.A. Lindqvist, G. Marklund, F.D. Wilder, D.T. Young, K. Torkar, J. Goldstein, J.C. Dorelli, L.A. Avanov, M. Oka, D.N. Baker, A.N. Jaynes, K.A. Goodrich, I.J. Cohen, D.L. Turner, J.F. Fennell, J.B. Blake, J. Clemons, M. Goldman, D. Newman, S.M. Petrinec, K.J. Trattner, B. Lavraud, P.H. Reiff, W. Baumjohann, W. Magnes, M. Steller, W. Lewis, Y. Saito, V. Coffey, M. Chandler, Science **352**, 6290 (2016). <https://doi.org/10.1126/science.aaf2939>

276. S.K. Godunov, Mat. Sb. **47**(3), 271 (1959). In Russian
277. B. van Leer, in *Lecture Notes in Physics*, vol. 18 (Springer, Berlin, 1973), pp. 163–168. <https://doi.org/10.1007/BFb0118673>
278. B. van Leer, J. Comput. Phys. **14**(4), 361 (1974). [https://doi.org/10.1016/0021-9991\(74\)90019-9](https://doi.org/10.1016/0021-9991(74)90019-9)
279. B. van Leer, J. Comput. Phys. **23**(3), 263 (1977). [https://doi.org/10.1016/0021-9991\(77\)90094-8](https://doi.org/10.1016/0021-9991(77)90094-8)
280. B. van Leer, J. Comput. Phys. **23**(3), 276 (1977). [https://doi.org/10.1016/0021-9991\(77\)90095-x](https://doi.org/10.1016/0021-9991(77)90095-x)
281. B. van Leer, J. Comput. Phys. **32**(1), 101 (1979). [https://doi.org/10.1016/0021-9991\(79\)90145-1](https://doi.org/10.1016/0021-9991(79)90145-1)
282. P.L. Roe, J. Comput. Phys. **43**, 357 (1981). <https://doi.org/10.1006/jcph.1997.5705>



Multiscale Kinetic Simulations

8

Giovanni Lapenta

Abstract

Plasmas are intrinsically multiscale. Electrons and ions respond to different scales, and usually macroscopic scales are much bigger than particle scales. Addressing this problem requires introducing adaptation in time, space, and phase space and using multiple mathematical models. We review here recent developments in using implicit and semi-implicit methods for handling multiple time scales. Spatial scales are addressed with space adaptive methods, r-adaptation to move grid points, and h-adaptation to introduce multiple levels of resolution using the multilevel-multidomain approach. Phase space is also adapted using methods to repopulate the particles. Finally, multiple levels of description, fluid and kinetic, are linked together.

8.1 Plasma Scales and Models

Plasmas are by their intrinsic nature multiscale and multi-physics [1]. Electrons and ions respond on very different scales by virtue of their different masses, and even electron-positron plasmas might have multiple masses (in the laboratory frame) when relativistic velocities are present and the range of Lorentz factors is wide. Plasmas have many waves and eigenmodes, ranging many orders of magnitude in space and in time. Figure 8.1 shows a typical example of the scales one would

G. Lapenta (✉)

Centre for Plasma Astrophysics, Department of Mathematics, Catholic University of Leuven, Heverlee, Belgium

e-mail: giovanni.lapenta@wis.kuleuven.be

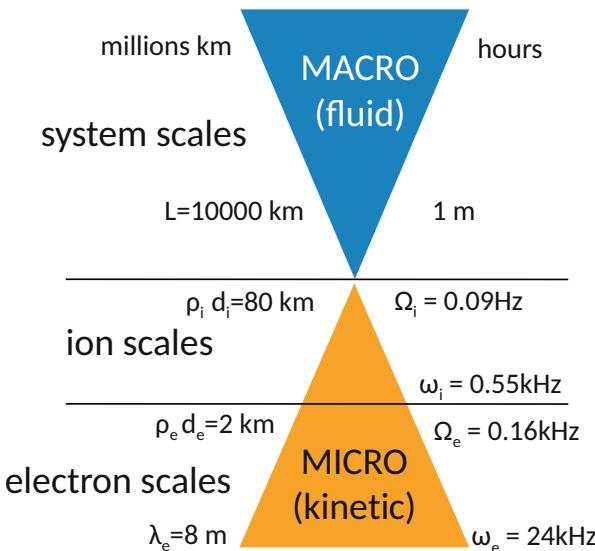


Fig. 8.1 Typical scales observed during missions of exploration in the Earth space environment. The spatial scales are on the left, and the temporal scales are on the right. Electrons are much lighter, and their scales are orders of magnitude smaller. A tremendous spread is present requiring advanced modeling techniques and the largest computing resources available

observe in the Earth space environment. For hydrogen plasmas, the mass ratio is 1836, and its square root determines the range of scales between electron and ion inertial lengths and plasma frequencies. The cyclotron frequencies differ by the mass ratio. The species beta β_s , the ratio of the species' pressure with the magnetic energy, can also vary over orders of magnitude in different space and astrophysical plasmas, determining the ratio between a species inertial length and its gyro-radius or its cyclotron and plasma frequency. In the case of the electrons, the Debye length is usually much smaller than the electron inertial length because of the low thermal speed compared with the speed of light: $\lambda_{De}/d_e = v_{the}/c$.

The disparity of scales poses a great challenge. If one is forced to resolve the smallest scales present in the plasma, only microscopic simulations would be possible. But luckily often, this level of resolution is not needed. For this reason, reduced models have been developed to capture the essential physics needed while excluding all unnecessary processes. The procedure is more suitable to cases where the understanding of the physics is well developed, but it becomes dangerous in cases of true exploration where the unknown needs to be probed: then only first-principle models can be truly trusted.

We consider here a different approach: instead of reducing analytically the model to design it to treat the specific processes of interest, we propose to use numerical adaptation. Numerical methods can resolve only the scales of interest if appropriate means are used to deal with the unresolved scales. The key element of the strategy is to use implicit methods. Implicit methods prevent unresolved scales from producing numerical instabilities that become a source of unphysical energy. Provided stability is ensured, using an implicit method is like using a reduced model that eliminates a physical process but with one critical difference: if one desires to reintroduce the process, one just need to increase the resolution. This approach then allows the user to conduct convergence studies varying the domain size and the time interval. It is an instrument of great power that needs to be used with experience, properly selecting the scales resolved.

The present chapter reviews our experience in designing and operating such implicit models. The assumption is made here that the reader is already familiar with the particle in cell (PIC) method. The present book provides a fine introduction, but interested readers are also pointed toward the recent introduction to particle in cell by the same author of the present chapter: the same notation used here facilitates following the arguments [2].

We address in Sect. 8.2 adaptation in time, considering implicit and explicit methods and their regimes of operation, focusing especially on the important concept of energy conservation. We consider then in Sect. 8.3 space adaptation considering both r-adaptation (moving grids with constant connectivity) and h-adaptivity (use of multiple interconnected levels of resolution such as adaptive mesh refinement).

Section 8.4 considers the peculiar need of adaptivity in particle in cell methods dealing with adaptation of particle description of phase space.

Finally, we consider the possibility of reintroducing in the approach reduced models by using physics adaptivity, with different processes described by different sets of equations.

8.2 Temporal Adaptation

Time is discretized in steps. We consider here some general properties applicable to a vast class of methods if not all. Time discretization can be either *explicit*, when one time step is directly computed from the previous time step without requiring iterations or solution of linear or nonlinear coupled sets of equations, or *implicit*, when the time step to be calculated appears in the calculation and an iteration (either linear or nonlinear) is needed. We review below the two methods and consider their advantages.

8.2.1 Implicit Versus Explicit

A general classical formulation of the discretization in time can be expressed as:

$$\begin{aligned}\mathbf{x}_p^{n_x} &= \mathbf{x}_p^{n_x-1} + \Delta t \langle \mathbf{v}_p \rangle_x \\ \mathbf{v}_p^{n_v} &= \mathbf{v}_p^{n_v-1} + \frac{q_p \Delta t}{m_p} (\mathbf{E}_p + \langle \mathbf{v}_p \rangle_v \times \mathbf{B}_p) \\ \mathbf{B}_g^{n_B} &= \mathbf{B}_g^{n_B-1} - \Delta t \nabla \times \langle \mathbf{E}_g \rangle \\ \mathbf{E}_g^{n_E} &= \mathbf{E}_g^{n_E-1} + c^2 \Delta t \nabla \times \langle \mathbf{B}_g \rangle - c^2 \mu_0 \langle \mathbf{J}_g \rangle\end{aligned}\tag{8.1}$$

Most of the existing scheme can be cast in this format with the appropriate choice of the temporal discretization levels n_x, n_v, n_B, n_E . The schemes are further defined by the choice of the averaged values of $\mathbf{E}_p, \mathbf{B}_p, \langle \mathbf{v}_p \rangle_x$ and $\langle \mathbf{v}_p \rangle_v$ averaged over the trajectory of a particle during the time step and by the averaged values of the fields and moments over the time step $\langle \mathbf{E}_g \rangle, \langle \mathbf{B}_g \rangle$ and $\langle \mathbf{J}_g \rangle$. The spatial discretization here is generic and indicated with the index g of the grid: this can be any of the schemes used in practice.

In principle, different numerical schemes can be designed making a specific choice for all these agents in Eq.(8.1). However, some fundamental physical conservation laws reduce the freedom in the design if the designer intends to enforce the conservation. We will consider the choices in turn. To express the time discretization, we introduce a subscript n_φ where φ specifies the time level of a specific quantity φ . Different quantities can have different time levels. The specific choice of where these levels are set defines the method used. Examples will be shown in Table 8.1.

First, a cardinal rule to follow is that the magnetic field should do no work. It is a fundamental property of magnetic fields; if we do not satisfy it, energy conservation will be lost, and particles will feel wrong magnetic forces and exchange energy incorrectly. No scheme known to the author does not conserve this basic sanity principle. To apply this requirement, we can use the second of Eq. (8.1) to compute the energy increase. As shown in Ref. [3], this can be obtained dotting this equation with $(\mathbf{v}_p^{n_v} + \mathbf{v}_p^{n_v-1})/2$ to obtain

$$\begin{aligned}\frac{m_p(v_p^{n_v})^2}{2} - \frac{m_p(v_p^{n_v-1})^2}{2} &= q_p \Delta t \mathbf{E}_p \cdot \frac{\mathbf{v}_p^{n_v} + \mathbf{v}_p^{n_v-1}}{2} \\ &\quad + \Delta t (\langle \mathbf{v}_p \rangle_v \times \mathbf{B}_p) \cdot \frac{\mathbf{v}_p^{n_v} + \mathbf{v}_p^{n_v-1}}{2}\end{aligned}\tag{8.2}$$

The specific choice of the dotting factor suggested [3] is critical because, thanks to the binomial product, it leads to energy conservation. Equation (8.2) expresses

the change in particle energy as a sum of two terms; one evaluates the work done by the electric field, while the last expresses the work done by the magnetic field. Obviously, the last term should be zero for any conceivable orientation and values of the magnetic field and particle velocities, a fact readily imposed by requiring that:

$$\langle \mathbf{v}_p \rangle_v = \frac{\mathbf{v}_p^{n_v} + \mathbf{v}_p^{n_v-1}}{2} \quad (8.3)$$

meaning that the speed used to compute the Lorentz force should be the algebraic mean of the velocities at the beginning and end of the time step for the velocity.

With this choice of $\langle \mathbf{v}_p \rangle_v$, the second of Eqs. (8.1) requires some algebraic manipulation to express the new velocity $\mathbf{v}_p^{n_v}$ that appears on both sides of the equation. This step does not require any solver and can be done analytically using the properties of the vector product. An often used algorithm is due to Boris and solves this equation in simple direct steps reviewed in any PIC textbook (see Chap. 6 and, e.g., [4–6]). However, to solve this equation, we need to have the average electric (\mathbf{E}_p) and magnetic (\mathbf{B}_p) fields acting on the particle along the trajectory followed in one time step.

This operation is a key definer of the specific scheme used. All schemes can be expressed using the interpolation function $S(\mathbf{x})$ to transfer information between particles and a grid used to compute the fields [2, 5, 7]. This step is perhaps the defining operation of the particle in cell scheme, and its complete use is reviewed in [2, 5–7]. Once an interpolation function is chosen, the electric and magnetic field can be expressed as:

$$\begin{aligned} \mathbf{E}_p &= \sum_g \langle \mathbf{E}_g \rangle S_E(\mathbf{x} - \langle \mathbf{x}_p \rangle_E) \\ \mathbf{B}_p &= \sum_g \langle \mathbf{B}_g \rangle S_B(\mathbf{x} - \langle \mathbf{x}_p \rangle_B) \end{aligned} \quad (8.4)$$

where the interpolation function for electric and magnetic field may in general be different. The fields are computed on the particles from the fields known on the grid using the interpolation functions. But even given the interpolation functions, different schemes can still be derived using different definitions of the time step averaging operator indicated by the overline in the definition of $\langle \mathbf{E}_g \rangle$ and $\langle \mathbf{B}_g \rangle$.

The opposite operation of going from the particles to the grid is required for computing the current:

$$\langle \mathbf{J}_g \rangle = \sum_p q_p \langle \mathbf{v}_p \rangle_J S_J(\mathbf{x} - \langle \mathbf{x}_p \rangle_J) \quad (8.5)$$

So far, we still retain great freedom in deciding many of the choices introduced above. Additional conservation laws further reduce this freedom. First, we consider the conservation of energy. Equation (8.2) expresses the change of the energy of a

particle. From the last two of Eq. (8.2), we can obtain the change of energy for the electromagnetic field. Dotting the third equation with $(\mathbf{B}_g^{n_B} + \mathbf{B}_g^{n_B-1})/2$ and the last with $(\mathbf{E}_g^{n_E} + \mathbf{E}_g^{n_E-1})/2$, we can obtain

$$\begin{aligned} & \sum_g \left(\frac{\epsilon_0(E_g^{n_E})^2}{2} + \frac{(B_g^{n_B})^2}{2\mu_0} - \frac{\epsilon_0(E_g^{n_E-1})^2}{2} - \frac{(B_g^{n_B-1})^2}{2\mu_0} \right) \\ &= \Delta t \sum_g \left(\nabla \cdot \mathbf{S}_g + \langle \mathbf{J}_g \rangle \cdot \frac{(\mathbf{E}_g^{n_E} + \mathbf{E}_g^{n_E-1})}{2} \right) \end{aligned} \quad (8.6)$$

The energy of the electromagnetic field is changed by two terms, and the first is written in summarized form as divergence of the Poynting flux discretized on the grid. This term holds its own complexities as the discretized operators need to have mimetic properties of the analytical operators to exchange the cross product with the curl operator [8]. We assume here that to be the case in the discretization scheme used. The last term expresses the work done by the electric field on the particles. Obviously, this term needs to be identical to that same term in the law of energy conservation for the particles, Eq. (8.2). Summing over all particles and grid cells to achieve overall conservation:

$$\sum_g \sum_p q_p \langle \mathbf{v}_p \rangle_J S_J (\mathbf{x} - \langle \mathbf{x}_p \rangle_J) \cdot \frac{(\mathbf{E}_g^{n_E} + \mathbf{E}_g^{n_E-1})}{2} = \sum_p \sum_g \langle \mathbf{E}_g \rangle S_E (\mathbf{x} - \langle \mathbf{x}_p \rangle_E) \cdot \frac{\mathbf{v}_p^{n_v} + \mathbf{v}_p^{n_v-1}}{2} \quad (8.7)$$

This equality has to be valid for any conceivable field and particle trajectory. A sufficient condition for energy conservation then requires that:

$$S_J (\mathbf{x} - \langle \mathbf{x}_p \rangle_J) = S_E (\mathbf{x} - \langle \mathbf{x}_p \rangle_E)$$

$$\begin{aligned} \langle \mathbf{v}_p \rangle_J &= \frac{\mathbf{v}_p^{n_v} + \mathbf{v}_p^{n_v-1}}{2} \\ \langle \mathbf{E}_g \rangle &= \frac{(\mathbf{E}_g^{n_E} + \mathbf{E}_g^{n_E-1})}{2} \end{aligned} \quad (8.8)$$

These requirements determine several design choices for a PIC scheme. The first and most important consequence of the energy conservation requirement is that explicit schemes cannot be used. Equation (8.8) requires that the electric field used to move the particles includes information from the new time level n_E , while the equation for the new electric field requires to know the new particle velocity. The field and particle equations cannot be solved independently. The scheme needs to be implicit.

Table 8.1 reports some of the most commonly used schemes in PIC simulation.

Table 8.1 Summary of the design choice of some of the most common algorithms used in particle in cell. Values identified by an overline are meant to be averages between the ends of the time step, and $\overline{\psi} = (\psi^{n+1} - \psi^n)/2$ and θ values are defined as $\psi^\theta = (\theta\psi^{n+1} - (1-\theta)\psi^n)$. θ is a decentering time parameter used by implicit and semi-implicit schemes [9, 10]. Exact energy conservation is achieved for $\theta = 1/2$

	Explicit	Implicit	Semi-Implicit (ECsim)
n_x	$n + 1$	$n + 1$	$n + 1/2$
n_v	$n + 1/2$	$n + 1$	$n + 1$
n_E	$n + 1/2$	$n + 1$	$n + 1$
n_B	$n + 1$	$n + 1$	$n + 1$
$\langle \mathbf{E}_g \rangle$	\mathbf{E}_g^n	$(\mathbf{E}_g^n + \mathbf{E}_g^{n+1})/2$	$(\mathbf{E}_g^n + \mathbf{E}_g^{n+1})/2$
$\langle \mathbf{B}_g \rangle$	$\mathbf{B}_g^{n+1/2}$	$(\mathbf{B}_g^n + \mathbf{B}_g^{n+1})/2$	$(\mathbf{B}_g^n + \mathbf{B}_g^{n+1})/2$
$\langle \mathbf{v}_p \rangle_x$	$\mathbf{v}_p^{n+1/2}$	$\overline{\mathbf{v}}_p = (\mathbf{v}_p^{n+1} + \mathbf{v}_p^n)/2$	\mathbf{v}_p^n
$\langle \mathbf{J}_g \rangle$	$\sum_p \mathbf{v}_p^{n+1/2} (S(\mathbf{x}_g - \mathbf{x}_p^{n+1}) + S(\mathbf{x}_g - \mathbf{x}_p^n))/2$	$\sum_p \overline{\mathbf{v}}_p S(\mathbf{x}_g - \overline{\mathbf{x}}_p)$	$\sum_p \overline{\mathbf{v}}_p S(\mathbf{x}_g - \mathbf{x}_p^{n+1/2})$
\mathbf{E}_p	$\sum_g \mathbf{E}_g^n S(\mathbf{x}_g - \mathbf{x}_p^n)$	$\sum_g \mathbf{E}_g^\theta S(\mathbf{x}_g - \overline{\mathbf{x}}_p)$	$\sum_g \mathbf{E}_g^\theta S(\mathbf{x}_g - \mathbf{x}_p^{n+1/2})$
\mathbf{B}_p	$\sum_g (\mathbf{B}_g^n + \mathbf{B}_g^{n+1}) S(\mathbf{x}_g - \mathbf{x}_p^n)/2$	$\sum_g \mathbf{B}_g^\theta S(\mathbf{x}_g - \overline{\mathbf{x}}_p)$	$\sum_g \mathbf{B}_g^\theta S(\mathbf{x}_g - \mathbf{x}_p^{n+1/2})$

In *explicit schemes*, the four Eq. (8.1) are solved in sequence, each being formulated only using quantities known from the previous time step, and no iteration is needed among the equations. The leap-frog algorithm for particles and fields does precisely that. The simplicity of the explicit scheme is extremely valuable in practical implementations, especially on massively parallel supercomputers where it turns out to be one of the most effective utilizers of the parallel architecture of any scheme used in computational science. The downside however is that explicit PIC does not conserve energy and has the most restrictive stability constraints making it least suitable to deal with multiple scale problems and adaptation.

In *implicit schemes*, instead, the information on the new time level derived for one of Eq. (8.1) enters another, and none can be solved independently. Obviously, the complexity of solving so many coupled equations (six equations per particles and six per grid location) is great. However, this is a type of complexity that is perfectly manageable in modern supercomputers using modern solvers.

Implicit methods further divide into two types. The first is *fully implicit* or *nonlinearly implicit*. In this case, the coupling between the Eq. (8.1) is nonlinear, and only using nonlinear solvers can deal with it: the fully implicit energy-conserving, ECPIC, scheme is one recently developed method of this type [3, 11]. The nonlinear coupling is highlighted in Table 8.1 by the definition for $\langle \mathbf{J}_g \rangle$ that requires to know the new particle velocity at the end of the time step, but to have it, one needs to first compute the particle electric field, which in turn requires to have solved the Maxwell equations that use the very $\langle \mathbf{J}_g \rangle$ we are trying to compute. The nonlinear coupling of the current with the electric field is the central motive for requiring a nonlinear iterative solver.

The second is *semi-implicit or linearly implicit*. In this case, instead, the Eq. (8.1) are formulated using the freedom we still have in the choices above, or making approximations, so that the remaining coupling is linear. The energy-conserving semi-implicit scheme, ECsim [10, 12–14], belongs to this category and uses a proper choice of the freedom still allowed in the formulation above to remain exactly energy conserving without any approximation. The design choice of moving the particles on a time staggered temporal marching scheme allows the position $\mathbf{x}_p^{n+1/2}$ to be known directly from the previous time step. This in turn makes all the interpolation functions used in the calculation of the moments on the grid and of the fields on the particles known. This removes the nonlinearity and makes the relationship between current and electric field linear. In the scheme of ECsim, this linearity is expressed via a *mass matrix* that summarizes the response of the particles to the fields[10].

Instead, the moment implicit [9] (e.g., iPic3D [15] and Celeste [16, 17]) and direct implicit [18] (e.g., Avanti [19]) methods introduce approximations to remove the nonlinearity [20] and in the process loose energy conservation. The downside is obvious, but it comes at a great simplification of the algorithm that no longer needs a nonlinear solver and consequently at a great reduction of the memory usage and CPU time required for each time step of the algorithm.

For example, the implicit moment method often used in space problems via the iPic3D code is based on using the first equation of motion in the definition of the current of the implicit model as

$$\langle \mathbf{J}_g \rangle = \sum_p q_p \bar{\mathbf{v}}_p S(\mathbf{x}_g - \mathbf{x}_p^n - \bar{\mathbf{v}}_p \Delta t / 2) \quad (8.9)$$

The interpolation function can then be expanded as [16]:

$$S(\mathbf{x}_g - \mathbf{x}_p^n - \bar{\mathbf{v}}_p \Delta t / 2) = S(\mathbf{x}_g - \mathbf{x}_p^n) - \nabla S|_{\mathbf{x}_g - \mathbf{x}_p^n} \cdot \frac{\bar{\mathbf{v}}_p \Delta t}{2} \quad (8.10)$$

The nonlinearity is eliminated and the current becomes:

$$\langle \mathbf{J}_g \rangle = \sum_p q_p \bar{\mathbf{v}}_p S(\mathbf{x}_g - \mathbf{x}_p^n) - \nabla \cdot \sum_p S(\mathbf{x}_g - \mathbf{x}_p^n) \frac{q_p \bar{\mathbf{v}}_p \bar{\mathbf{v}}_p \Delta t}{2} \quad (8.11)$$

The role of the last term (related to the pressure tensor but defined with particle charges and not masses) becomes clear, showing how the implicit moment method uses higher-order moments to compute the current, giving the method its name. Making the Taylor series expansion truncated to first order to eliminate the nonlinearity in the expression of the current allows the use of linear solver but breaks energy conservation. The linear iteration is obtained solving analytically the equation of the particle momentum, substituting it in Eq (8.11) to obtain a linear relationship between current and electric field [2, 7, 16, 17].

In essence, the explicit method does not conserve energy but at its core does not require any linear solver or iterations, the implicit method conserves energy

exactly, but it requires a nonlinear solver. The semi-implicit method in its most recent ECSim implementation achieves both goals of conserving energy and not requiring a nonlinear solver. The semi-implicit method still needs global linear solver, a requirement that has significant consequences for the scalability of massive parallel implementations.

In the discussion above, the focus was on energy conservation; other conservation laws have not been mentioned. Momentum conservation is sought to be incompatible with energy conservation: it appears numerical methods can preserve one but only at the expense of the other [21]. It is, however, possible to conserve momentum if energy is sacrificed [6]. It is also possible to modify the current accumulation formula to impose detailed charge conservation and conservation of the Gauss theorem. PIC methods, of course, conserve charge globally because the charge of the particles is invariant. However, when solving the two curl equations of Maxwell's equations numerically, it is not assured that the two divergence constraints are still satisfied. Conserving the null divergence of the magnetic field is enforced by requiring that the divergence of the curl operator is still zero in the discretized scheme. But imposing Gauss theorem requires more than just the vanishing of the curl of the electric field has zero divergence, and it also requires that the divergence of the current satisfies the continuity equation for the charge:

$$\rho_g^{n+1} = \rho_g^n - \Delta t \nabla \cdot \langle \mathbf{J}_g \rangle \quad (8.12)$$

This equation can be enforced in both explicit [22–24] and implicit schemes [25] to obtain exact charge conservation.

8.2.2 Stability of Explicit and Implicit Schemes

The difference between explicit and implicit schemes has another important aspect: numerical stability. Explicit schemes are simpler and more efficient on parallel implementations, but implicit schemes can still win when stability is considered.

Stability of PIC algorithms needs to consider three aspects: the stability due to the discretization of the Maxwell equations in time, of the Vlasov equation in time, and to aliasing. We will consider them in turn.

Considering the electromagnetic case introduces the need to discretize the temporal variation of the electromagnetic field. This operation can itself also be done explicitly or implicitly. The discretization of the Maxwell's equations is well covered elsewhere [26], and here we just limit ourselves to say that if such discretization is explicit, the CFL (Courant-Friedrichs-Lowy) condition on the speed of light needs to be satisfied, $c\Delta t < \zeta \Delta x$, with some factor ζ of order unity, but not necessarily identical to one (it could be $1/2$, $1/\sqrt{3}$, $1/\sqrt{2}$ or some other number depending on the exact scheme chosen). Implicit discretization of the Maxwell's equation instead is unconditionally stable. It is possible to design explicit PIC algorithms that are implicit only in the field equations [27]: these are still explicit in the particle-field coupling, and the implicitness is present only in solving Maxwell's

equations. A linear solver is then needed only for the field equations. The advantage of this approach is removing the speed of light CFL conditions, but other plasma scales still need to be resolved for stability.

The stability conditions specific to particle in cell methods, instead, can be best understood in the case of an electrostatic plasma composed of electrons and immobile ions where the Vlasov-Poisson system can be linearized and solved assuming a Fourier transformation in space to obtain:

$$\frac{\partial \tilde{f}(\mathbf{k}, \mathbf{v}, t)}{\partial t} = \mathcal{H}(\mathbf{k}, \mathbf{v}) \tilde{f}(\mathbf{k}, \mathbf{v}, t) \quad (8.13)$$

where $\tilde{f}(\mathbf{k}, \mathbf{v}, t)$ is the Fourier transforms of the perturbation of the electron distribution function. The integro-differential operator $\mathcal{H}(\mathbf{k}, \mathbf{v})$ represents the linearized Vlasov-Poisson operator in Fourier transformed space. Its solution with the Laplace transformation leads to the Landau contour and the dispersion relation that gives all the plasma normal modes [28, 29].

The different behavior of implicit and explicit schemes can be understood by discretizing the linearized equation as [30]:

$$\tilde{f}^{n+1}(\mathbf{k}, \mathbf{v}) - \tilde{f}^n(\mathbf{k}, \mathbf{v}) = \Delta t \mathcal{H}^{n+\theta}(\mathbf{k}, \mathbf{v}) \quad (8.14)$$

A θ -scheme is used with $\mathcal{H}^{n+\theta} = \theta \mathcal{H}^{n+1} + (1 - \theta) \mathcal{H}^n$. The scheme is unstable for $\theta < 0.5$, Crank-Nicholson implicit when $\theta = 0.5$, and implicit Euler when $\theta = 1.0$. All values $\theta \geq 0.5$ result in unconditionally stable discretizations. The Von Neumann stability analysis confirms these properties for the scheme in Eq. (8.14). We assume a dependence of

$$\tilde{f}^n(\mathbf{k}, \mathbf{v}) = \hat{f}(\mathbf{k}, \mathbf{v}) e^{-i\omega n \Delta t} \quad (8.15)$$

and

$$\tilde{\mathcal{H}}^n(\mathbf{k}, \mathbf{v}) = \hat{\mathcal{H}}(\mathbf{k}, \mathbf{v}) e^{-i\omega n \Delta t} \quad (8.16)$$

We can cast Eq. (8.14) in the form:

$$\frac{e^{-i\omega \Delta t} - 1}{\theta e^{-i\omega \Delta t} + 1 - \theta} \hat{f}(\mathbf{k}, \mathbf{v}) = \Delta t \hat{\mathcal{H}}(\mathbf{k}, \mathbf{v}) \quad (8.17)$$

If we consider the exact linear Vlasov-Poisson equation without temporal discretization (8.13), assuming harmonic dependence

$$\tilde{f}(\mathbf{k}, \mathbf{v}, t) = \hat{f}(\mathbf{k}, \mathbf{v}) e^{-i\Omega t} \quad (8.18)$$

it follows that:

$$-i\Omega \hat{f}(\mathbf{k}, \mathbf{v}) = \hat{\mathcal{H}}(\mathbf{k}, \mathbf{v}) \quad (8.19)$$

There is then a simple relationship of the numerical discretized solution of the linear Vlasov equation with the exact theoretical solution:

$$\frac{e^{-i\omega\Delta t} - 1}{\theta e^{-i\omega\Delta t} + 1 - \theta} = -i\Omega\Delta t \quad (8.20)$$

Figure 8.2 shows the relationship between the plasma normal modes seen by the discretized Vlasov equation, ω , and the normal modes seen by the exact Vlasov equation, Ω . We assume here the plasma mode to be purely real, the case of an undamped wave. As can be seen, a value of $\theta = 0.5$ maintains stably any waves regarding its natural frequency. However, the wave frequency is distorted and lowered, limiting it to the Nyquist frequency $\pi/\Delta t$. For any value of $\theta > 0.5$, waves are not only limited in frequency but are also damped.

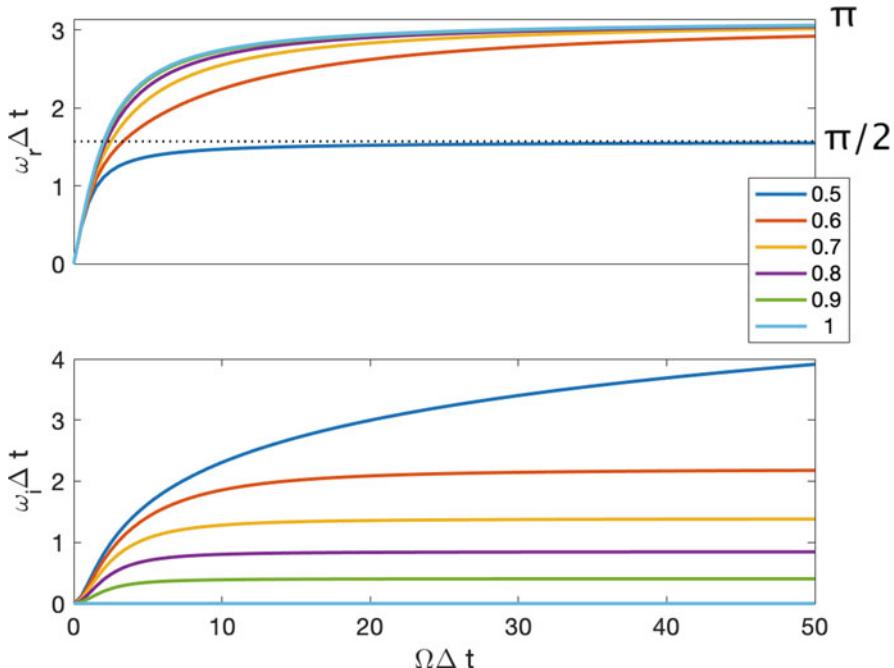


Fig. 8.2 Relationship in Eq. (8.20) between the plasma normal modes seen by the discretized Vlasov equation, ω , and the normal modes seen by the exact Vlasov equation, Ω , assumed real. Six different values of θ are reported with different colors indicated in the legend. The real part is above, the imaginary part below. The limit values of $\omega_r \Delta t = \pi/2, \pi$; are identified on the right axis

The practical result is that unresolved waves with frequencies that the time step cannot resolve are present in the discretized system but are damped, with a damping increasing with θ and with the time step.

The same analysis can be carried out for the leap-frog scheme used in explicit PIC:

$$\tilde{f}^{n+1/2}(\mathbf{k}, \mathbf{v}) - \tilde{f}^{n-1/2}(\mathbf{k}, \mathbf{v}) = \Delta t \mathcal{H}^n(\mathbf{k}, \mathbf{v}) \quad (8.21)$$

giving the famous formula $\sin(\omega \Delta t / 2) = \Omega \Delta t / 2$. Since \sin of real arguments is bound between -1 and 1 , the equation gives complex solutions, for $\Omega \Delta t > 2$. Figure 8.3 shows the solutions. Above the limit, a positive imaginary part is produced, leading to an amplification of the normal mode in time. This condition is called numerical instability and represents a physical error, an error that produces numerical sources of energy for the plasma. A cardinal rule in PIC is that this instability cannot be accepted, requesting $\Omega \Delta t < 2$. In practice, explicit schemes require even more stringent conditions to avoid severe energy nonconservation. Typically, $\Omega \Delta t < 0.1 - 0.5$ is required, with Ω being the fastest mode present in the plasma. The practical result is that simulations must select the time step to resolve the fastest frequency in the system, even in cases when that frequency carries no interesting physics. In multiscale processes, this limitation makes it impossible to model longer time scale processes. In these cases, the implicit method is preferable because it allows

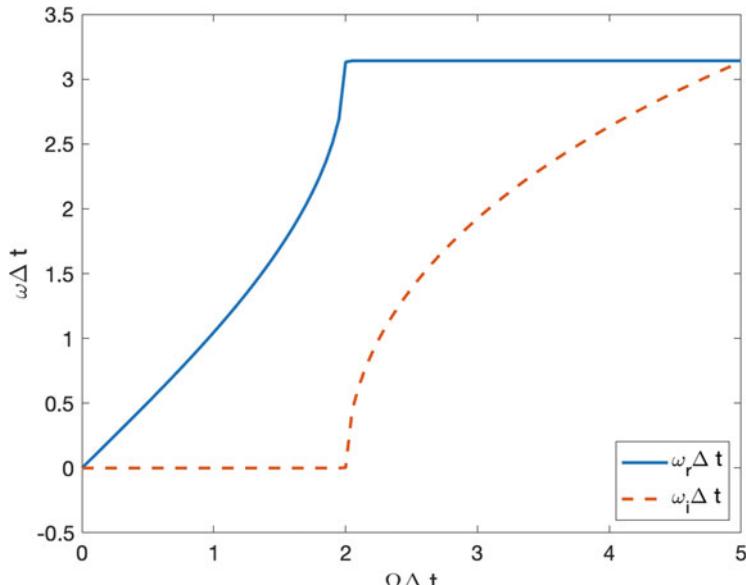


Fig. 8.3 Relationship for the explicit leap-frog method between the plasma normal modes seen by the discretized Vlasov equation, ω , and the normal modes seen by the exact Vlasov equation, Ω , assumed real. The real part (solid line) and imaginary part (dashed) are reported

to adapt the simulation to the temporal scales of interest, without the unresolved scales producing unphysical energy sources due to numerical instabilities.

But there should be clarity, the implicit approach resolves all scales resolved by the Δt , and it is the user who decides what scales to resolve. With the implicit method the user can select to resolve even the fastest scales in the system, there is no limitation. This is a great advantage over reduced methods that actually completely remove selected physics processes.

The analysis above has assumed space to be continuous, but of course, PIC methods use a grid to discretize the field equations. This is the source of the *finite grid instability*: the particles move in a continuum, but the fields are computed on the grid. The grid can only describe a finite number of modes, but the particles can describe any continuum wave number. Aliasing gives rise to an instability whose specific properties can be described by a detailed analysis of the evolution of Fourier modes. The analysis is mathematically involved, especially when time and space discretization are considered concurrently. Additionally, the specific choice of interpolation function used in the scheme affects the stability properties. The reader is referred to [4, 5] for a detailed analysis. As a rule of thumb, explicit methods need to resolve a multiple of the Debye length for stability, $\Delta x/\lambda_{De} < \xi$, where ξ depends on the details of the scheme used. As a consequence, explicit methods are forced to resolve all scales in the system: they need the time step to resolve the fastest frequency, $\Omega\Delta t < 2$, and the grid spacing to resolve (a small multiple of) the Debye length. Additionally, the two need to satisfy the CFL condition, $c\Delta t < \zeta\Delta x$, if the method is also explicit in the discretization of the Maxwell equations.

Implicit methods eliminate completely the time step and CFL condition, but they can still suffer from the aliasing instability. However, the condition is much relaxed. To understand the reason, we can recall that $\lambda_{De} = v_{the}/\omega_{pe}$. The stability condition can then be recast as

$$\frac{\Delta x \omega_{pe} \Delta t}{v_{the} \Delta t} < \xi \quad (8.22)$$

In the analysis above for implicit methods, we have observed how for large time steps the real frequency of unresolved modes is limited by the Nyquist frequency, $\omega_{pe}\Delta t \rightarrow \pi$, and the stability constraint for implicit methods to aliasing instability becomes:

$$\frac{v_{the} \Delta t}{\Delta x} > \eta \quad (8.23)$$

with η depending on the scheme used. The value of η is usually very small allowing a large working range than used together with the mover accuracy limitation. This result means that grid spacing cannot be much smaller than the average distance made by a particle in one time step. This is a very counterintuitive requirement: the time step cannot be too small. If one refines the time step without refining space also, the implicit scheme becomes unstable. This is the reason why in Fig. 8.1, we have introduced the hourglass. Time step and grid spacing can be chosen coarse or

fine, the user is free to select the range desired, but they must be both small or both large. Space and time need to be chosen together.

The value of η is much harder to pinpoint than that of ξ in explicit schemes, and it is not only scheme dependent but problem dependent: experimentation on a new problem is required to gauge the actual limit. Energy-conserving schemes have a much broader stability limit than other implicit schemes: energy is bound to remain constant, and energy is not created by the numerical instability. However, momentum is not conserved and can be distorted by the aliasing instability causing severe problems. Among uniform plasmas, energy-conserving implicit schemes are unconditionally stable and do not suffer from any aliasing instability (except in presence of an average electron flow but average flows can always be eliminated by a frame change, for uniform not sheared flows). The study of stability to the finite grid instability of implicit schemes in nonuniform plasmas is an open question for future research.

It should be pointed out that the instabilities above always affect the electrons and not the ions, so all considerations above are relative to electrons and the analysis of stability can be done assuming immobile infinitely massive ions.

Stability is not the only concern. Accuracy decreases when a particle travels considerable distances in one time step. The mover assumes a linear trajectory and uses a single force computed at a specified location: the motion of a particle becomes inaccurate if the linear straight trajectory is too long. Additionally, plasmas tend to satisfy the important frozen-in condition, so if particles move significantly in one time step, fields would do too. As a consequence, accuracy requires particles not to move too much in one time step. The condition $v_{the}\Delta t < \Delta x$ is often quoted [9] to guarantee that the particle motion in one time step does not overstep one cell to ensure the force acting on the particle is accurate. If a particle were to move more cells but the force were computed from only one of them, the trajectory would be inaccurate. Subcycling the particle motion can ameliorate the situation [11], especially for very energetic particle, but even subcycling cannot ensure the frozen-in condition on the fields. Accuracy requires a condition of the type above to be satisfied, but its precise limit is problem and scheme dependent.

8.3 Spatial Adaptation

PIC have some features requiring special attention when deploying space adaptation. There is a fundamental reason: particles have a finite shape, and if this finite shape is not constant, particles can experience self-forces and momentum nonconservation [31]. PIC adaptation then requires special care to avoid unphysical processes in regions of changing grid resolution. Nevertheless, very often in space plasmas, the region of interest is localized but embedded in a wider domain (the universe) and spatial adaptation is very attractive.

PIC can be adapted in two ways, like many other methods. First, *r-adaptation* is based on moving grid points to focus the interest in regions of interest. In this approach, the grid remains logically cartesian with a connectivity that still

resembles that of a uniform grid. Second, *h-adaptation* is based on refining grid using multiple levels of resolution, as in the famous adaptive mesh refinement (AMR) approach [32]. A third possibility has also been successfully implemented: the use of unstructured meshes where the grid can be adapted at will and conform to complex geometrical objects [33]. This approach has the freedom of selecting the grid with great generality but has the geometrical complexity of overlapping particles of finite shape over irregular cells: this is a mundane task but one that requires careful implementation.

8.3.1 PIC with r-Adaptation

R-adaptation or moving mesh adaptation uses techniques of mesh generation to distribute the points of the grid in regions of interest. The grid can evolve in time to follow moving features if necessary.

Grid adaptation is a vast domain of computational physics, and the reader is referred to the specialized literature[34]. The fundamental idea is that of creating a mapping, which is a transformation between a logical, ξ , and a physical space, x . In the logical space, the grid is uniform and cartesian $\xi_{ijk} = i\xi_1 + j\xi_2 + k\xi_3$ where ξ_i is the natural basis in logical space. The mapping from the logical to the physical space is $x(\xi_{ijk})$ and determines the physical location of each grid point. Figure 8.4 illustrates the concept. The physical grid can be nonuniform, and the physical domain can conform to features. The logical grid, instead, remains uniform and rectangular.

The task of designing an r-adaptive PIC includes defining a grid generation scheme, moving particles in the physical and logical domain, solving field equations in a nonuniform grid, and interpolating between particles and fields.

Generating a grid with specific target properties is a topic attracting great attention and one of the fundamental tasks of computational science [34]. In application to PIC, one typically defines a merit function that is equidistributed on the grid: where the merit function is large, the cells are fine, and where it is small, the cells are coarse. In essence, the grid distributes the points focusing in the regions of high merit function. The choice of the merit function and of the method to generate the grid from it has received attention within PIC applications [35–38].

The merit function can be chosen to represent geometric features such as objects whose boundaries need to be well resolved or using physical quantities indicative of importance in the physical processes (e.g., currents). Once the merit function is defined, distributing it can be done, minimizing an equidistribution principle that leads to Euler-Lagrange equations [38, 39]. Alternatively, a moving mesh approach can be used to move the grid points to equidistribute the merit function [40].

Once the grid is generated, the metric tensor of the transformation between the logical and physical grid is known, and all other operations can be designed [34]. The field equations can be discretized on the nonuniform grid, for example, using finite volume [41, 42], finite difference schemes [43], or discrete exterior calculus form of electromagnetism[44]. Interpolations between particles and cells are computed in the logical grid where the usual cartesian formulas can be used,

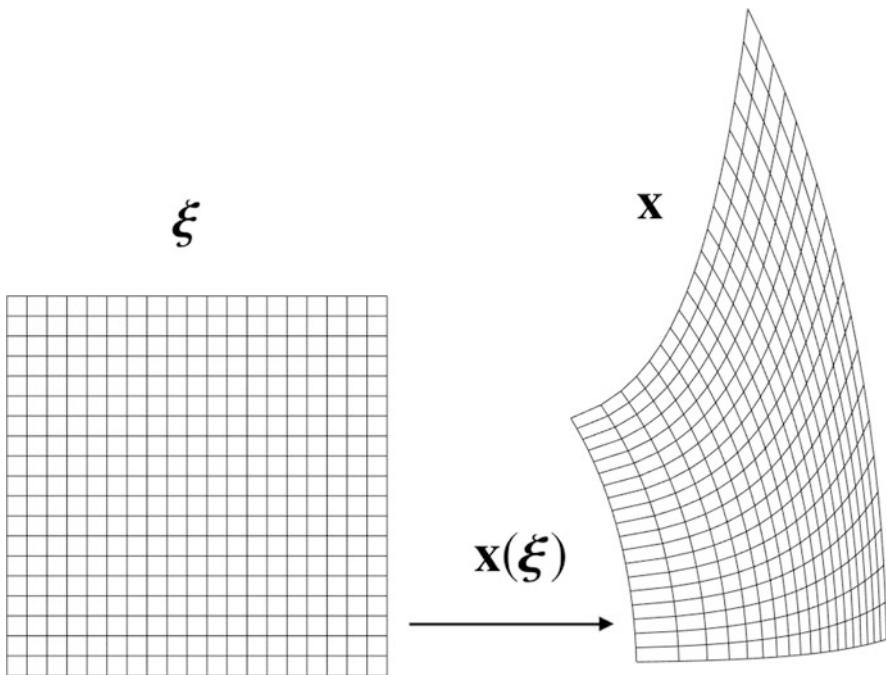


Fig. 8.4 Mapping $x(\xi)$ from the logical ξ to the physical x grid

typically using a b-spline of prescribed order as in the usual uniform grid PIC [41]. Note, however, that in a nonuniform grid, this means the particle shape and interpolation functions when seen in the physical grid change as the particle move: the particle shape is invariant only in the logical domain. This results in artificial effects and in particular to self-force and momentum nonconservation [45].

Particle motion in nonuniform grids can be done in principle in three ways: the equations of motion can be solved in the physical domain, in the logical domain, or with a mixed approach where the particle position is in the logical coordinates but the velocity is physical. Writing the particle equations of motion fully in the logical domain is complex, leading to inertial forces due to the coordinate transformation, and this approach is not typically followed. Solving the equations of motion fully in the physical domain is accurate, but it requires at the end of each time step to find where the particles are in the logical domain [46]. This operation requires to invert for each particle the coordinate transformation, determining the logical coordinate: $\xi_p = \xi(x_p)$. When the transformation is analytical, the task can be simple, provided the equations of the transformation can be analytically inverted; otherwise, they require a nonlinear search for each particle. A good compromise is to solve the equation for the momentum of the particles in physical space, but the equation for the particle position is the logical space [47]. In this case, the equation becomes:

$$\frac{d\xi_p}{dt} = \nabla\xi \cdot \mathbf{v}_p \quad (8.24)$$

In this case, the logical position is computed directly and no localization algorithm is needed. However, the logical position is correct to the truncation error used in the discretization of the Jacobian of the coordinate transformation $\nabla\xi$.

8.3.2 PIC with h-Adaptation

In other areas of computational physics, h-adaptation has obtained stellar successes[32]. In related fields of fluid plasma modeling, h-adaptation has been successfully applied [48–50]. In this approach, one introduces more points of description refining the mesh adaptively by further subdividing cells and increasing the resolution. The adaptive mesh refinement (AMR) is illustrated in Fig. 8.5. A domain is identified as requiring a higher resolution, and the local cells are further subdivided. There are many approaches to handle the complexity of the connectivity between the points in the different levels of description as well as to discretize the field equations on cells bordering jumps in resolution.

We focus here specifically on the application to PIC. In PIC, the grid is only one of the two main actors, the particles being the other. When particles cross a region of different resolution, the interpolation properties change: in essence, the computational particle changes shape. This effect results in a self-force, and special

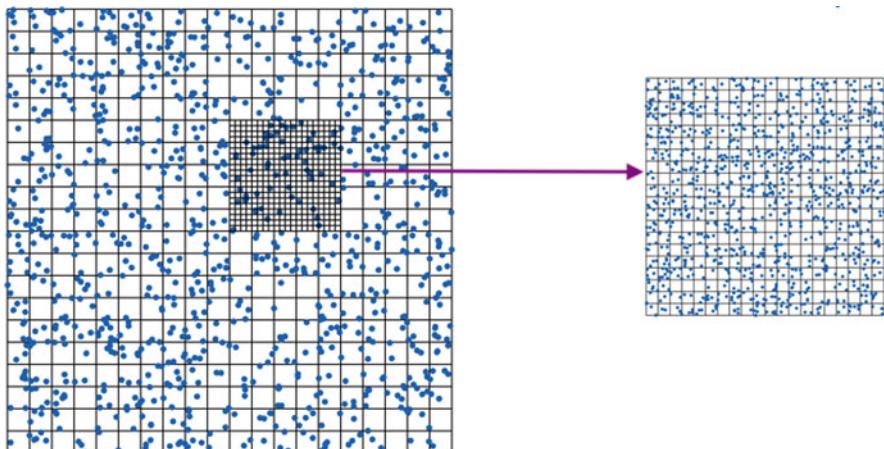


Fig. 8.5 Adaptive mesh strategies in PIC. On the left, the traditional adaptive mesh refinement (AMR) approach inserts a level of higher resolution for the grid, while the particles move between the two levels of resolution. On the right, the multilevel-multidomain (MLMD) approach describes the region of desired higher resolution as a new domain that exchanges information with the coarser level of resolution. Each domain has its own particles and fields, being conceptually a self-consistent complete model. Information between the two domains at the two different level is exchanged to keep consistency

methods have been developed to ameliorate this problem [31, 51, 52]. Additionally, the regions of smaller cells would tend to have fewer particles per cell. This is not desirable as the regions of higher resolution are the regions of greatest interest, and one would want more particles there. For this reason, AMR methods are usually coupled with the particle adaptation methods described in the next section 8.4. Despite these problems, successful AMR-PIC methods have been developed [52–54].

We have instead followed a different approach, shown on the right panel of Fig. 8.5: the multilevel-multidomain (MLMD) approach [55, 56]. In this approach, the region of higher resolution is spun off as a completely new daughter domain treated in principle by the same PIC method used in the mother domain. Each domain runs exactly the same operations, a great advantage in object-oriented massively parallel implementations because the same objects and same methods are called for all domains [57]. Simply, the scale of each domain and its physical location is different. In principle, the process can be repeated for different daughter domains in different regions of the mother domain or different generations of daughter domains with progressively higher resolutions can be used. The daughter domains can also move and even expand [58] and are not bound to the same location of the mother domain where they were generated.

There is some complexity in communicating information among the domains. The mother domain is used by the daughter domain to establish boundary conditions, for the fields but also for the particles entering the daughter domain [55, 56]. This step has some of the same complexity of the AMR approach, and it is addressed with the repopulation region. Each daughter has a buffer region of boundary cells where the particles are repopulated from information from the mother domain. The number of particles repopulation is selected to keep the desired number of particles per cell.

Conversely, the mother domain gathers from the daughter the more refined and accurate field of the daughter domain to improve the field information. The mother fields are not simply replaced. We have shown a better momentum conservation when the average between mother and daughter fields are used [59, 60].

The communication exchange is straightforward, but it requires a proper implementation within a parallel implementation [57]. Performance issues need to be considered carefully. In our implementation, each domain has its own processors to perform the operation: communication between the domain is implemented via MPI messages. The ordering of operations is designed to minimize the idle time when a domain waits for the operations done on the other, but some idle time is unavoidable. For this reason, using different time steps on the different domains might sound appealing because finer domains need finer time steps that are not necessary for the coarser mother domain [61]. However, this approach may incur in more idle time on the mother processors, and the issues raises complex computer science issues because idle time means also reduced energy consumption [61]. These problems are not peculiar of MLMD but general to all AMR approaches and to a lesser degree to r-adaptation wherein presence of large changes in grid spacing might also invite

using local time stepping, rising complex optimization issues in massively parallel implementation.

For best parallel performances, nothing beats a simple uniform explicit PIC with no global communication and only local processor-to-processor communication. But the advantages of the higher complexity of adaptive or implicit method can far outweigh the implementation difficulties [62].

8.4 Phase Space Adaptation

Particle adaptation is needed to increase the number of particles in regions where high accuracy is required and to reduce the number of particles where lower accuracy can be tolerated [63]. The primary effect of increasing the number of particles is to reduce the variance of the statistical description of the distribution function, a technique widely used in statistical and Monte Carlo methods [64].

Particle adaptation must be in effect during the calculation to keep the local required accuracy. When used in conjunction with adaptive grids, particle adaptation tries to keep the number of particles per cell constant. This approach leads to finer grid spacing in the region of interest and, automatically, to a higher density of computational particles in that region.

Adapting particles requires to enforce two different conservation properties [63, 65]. First, particle-grid interpolation needs to be conserved. This requirement imposes that the new set and old set of particles must have the same moments on the grid:

$$\sum_{N_p} S(\mathbf{x}_g = \mathbf{x}_p) q_p u_p^m = \sum_{N_{p'}} S(\mathbf{x}_g = \mathbf{x}_{p'}) q_{p'} u_{p'}^m \quad (8.25)$$

The new set indicated with prime quantities is composed of $N_{p'}$ particles and must respect the moments of the original set of N_p particles. There are, of course, an infinite number of moments, but most PIC schemes use only a finite set, typically only the moments 0 (density), 1 (current), and sometimes the pressure tensor (order 2). Second, the statistical distribution of the original set of particles and that of the new set must be compatible. Statistically, they have to be two different realizations of the same statistical distribution.

There are two approaches to imposing this requirement. The first is to replace a set of a few particles with a new set with a different, larger or smaller, number. The number of constraints, even assuming a number of moments up to order 2, is typically underdetermined because there are many more particles per cell than moments per cell to impose. A minimization procedure can then be used to identify a unique solution [66, 67].

One approach is to impose desirable numerical properties, such as the uniformity of the distribution of particles per cell or minimizing the entropy of the particle distribution [66]. Another approach is to use dimensionality reduction methods of machine learning to replace a set of particle with another using the k-means

approach [67, 68]. These approaches tend to be approximate and conserve the distribution and its moments within a certain tolerance of error.

Alternatively, a more exact approach is to rely on the general properties of the particle interpolation functions $S(\mathbf{x}_g = \mathbf{x}_p)$ to replace one particle with many. In the specific case of b-splines, a b-spline with support Δx can be replaced with multiple b-splines of fractional support [63, 69]. This general property of b-splines allows to impose all moments to be conserved but is limited to specific ratios in the number of particles of the new and old set as well as specific locations for the new particles with respect to the old. The advantage is that the method is exact and does not perturb the statistical distribution, for the operation of splitting one particle into many all with the same velocity. The inverse operation is in principle possible, but it would require to find particles in the right position with respect to each other and with the same velocity. Obviously, this is not possible, and an algorithm must be developed to approximate this operation. Below, we describe our algorithm [65] used and modified often in the literature [70–75].

The central idea is to reduce the variance (noise) of the simulation by splitting selectively computational particles into new particles. The properties of the new particles are chosen in order to preserve the relevant properties of the system. In a PIC plasma simulation, these properties are:

- local particle velocity distribution
- macroscopic moments measured on the grid (e.g., density, momentum, and pressure)

The inverse operation can be used where too many particles are present and the number of particles can be safely reduced. A method for particle control has been developed to account for these two requirements. The results derived in our previous work [63, 65] are summarized here.

For increasing the number of particles, a splitting algorithm is used:

Splitting Algorithm

Given a cell with N_p particles in a 1D, 2D, or 3D system, any chosen particle (labeled o) with charge q_o (and mass obtained from the charge-to-mass ratio for the species), position \mathbf{x}_o (in logical coordinates), and velocity \mathbf{v}_o can be replaced by N' particles, labeled $p' = \{1, 2 \dots N'\}$. In 1D, $N' = 2$ and the new properties are $q_{p'} = q_o/2$, $x_{p'} = x_o \pm 1/N_p$ (where the cell size is unitary), $\mathbf{v}_{p'} = \mathbf{v}_o$. In 2D, $N' = 4$ and the new properties are $q_{p'} = q_o/4$; $x_{1,2} = x_o \pm 1/N_p$, $x_{3,4} = x_o$, $y_{1,2} = y_o$, $y_{3,4} = y_o \pm 1/N_p$; $\mathbf{v}_{p'} = \mathbf{v}_o$. In 3D, $N' = 6$ and the new properties are $q_{p'} = q_o/6$, $x_{1,2} = x_o \pm 1/N_p$, $x_{3,\dots,6} = x_o$; $y_{1,2,5,6} = y_o$, $y_{3,4} = y_o \pm 1/N_p$; $z_{1,\dots,4} = z_o$; $z_{5,6} = z_o \pm 1/N_p$.

Note that the choice of the particle $p = o$ in the set of N_p particles in the cell is free. In the result sections, we choose the particle with the largest energy: $m_p \mathbf{v}_p^2$.

To decrease the number of particles, the splitting operation can be inverted to coalesce two particles into one. The difficulty is that, in general, it is impossible to find two particles with the same velocity. For this reason, particles with different

velocity have to be coalesced. To minimize the perturbation of the velocity distribution, the particles to be coalesced must be chosen with similar velocity. In practice, the following coalescence algorithm is used:

Coalescence Algorithm

Given a cell with N_p particles in 1D, 2D, or 3D systems, choose $N = 2$ particles $p = \{1, 2\}$ close to each other in the phase space. Their properties are q_p , \mathbf{x}_p , and \mathbf{v}_p . The two chosen particles can be replaced by one particle (labeled A) with $q_A = q_1 + q_2$, $\mathbf{x}_A = (q_1 \mathbf{x}_1 + q_2 \mathbf{x}_2)/q_A$, $\mathbf{v}_A = (q_1 \mathbf{v}_1 + q_2 \mathbf{v}_2)/q_A$.

The coalescence algorithm preserves the charge and momentum but perturbs the velocity distribution. The crucial point of the algorithm above is to choose two particles close in velocity and space. A pair search of the two particles closest in velocity is usually too expensive. For this reason, we perform a dichotomic search that sorts the particles into two bins and selects the largest bin. The binning is repeated in sequence for each spatial direction and component of the velocity. The binning is continued until the number of particles in the largest bin is small enough to use a pair search. More advanced coalescence algorithms can preserve more closely the conservation properties [67, 68, 74, 75].

8.5 Model Adaptation

Different descriptions of plasma can be obtained with the moment approach. The moments are defined as integrals of the distribution function multiplied by a power of the velocity [76]: density, velocity, pressure tensor, and heat flux are example of moments that can be derived. Starting from the Boltzmann equation, the moment approach produces fluid models based on the moments [1]. However, the equation for each moment requires moments of higher orders [77]. The central task of any fluid model is how to decide the closure, that is, how to approximate the higher-order moments using the moments available. Additionally, in applications seeking macroscopic behavior, often the plasma is treated as a single fluid, the well-known magnetohydrodynamics (MHD) method [78].

Therefore, fluid and kinetic formulations are not incompatible. Fluid models can be rigorously derived from kinetic models using the moment approach. However, the reverse is not true because by taking the moments, information on the phase space distribution is lost. When exchanging information, it is then obvious how to transfer information from the kinetic model to the fluid model, but the reverse requires additional considerations.

Within the MLMD approach, we intend to generate a kinetic domain interconnected with the fluid model. Figure 8.6 summarizes the idea. At the fluid level, the whole system is covered while the PIC level covers only a subdomain.

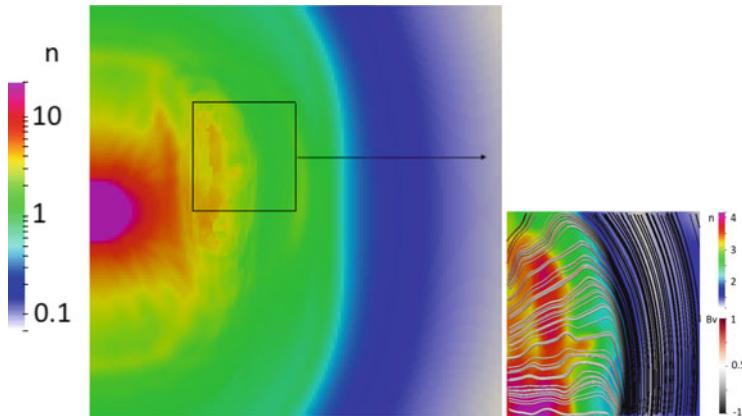


Fig. 8.6 Illustration of the coupling of a fluid and a kinetic domain. The information needed to initialize the kinetic domain is obtained from the MHD domain. The left domain is 2AU in size and shows a CME ejected from the Sun studied with the Slurm MHD code[79]. A subdomain identified by a box is used to create a kinetic simulation using the methods described in the text

8.5.1 Linking Kinetic and Fluid States

A kinetic state is not fully defined by a fluid state. The electric and magnetic fields are given, but the particle information is provided only via the first moments: density, velocity, and temperature. This information cannot determine a kinetic distribution function. An arbitrary assumption is needed. An obvious choice is the Boltzmann distribution because the H theorem states that in all systems, entropy tends to increase, and at equilibrium, it reaches a maximum [80] where the drifting Maxwellian distribution applies:

$$f_s(\mathbf{x}, \mathbf{v}, t) = \left(\frac{m_s}{2\pi k T_s(\mathbf{x}, t)} \right)^{3/2} n_s(\mathbf{x}, t) \exp \left(-\frac{m_s (\mathbf{v} - \mathbf{v}_s(\mathbf{x}, t))^2}{2k T_s(\mathbf{x}, t)} \right) \quad (8.26)$$

for each species s (electrons or ions). However, space plasmas are typically not in equilibrium, and nonequilibrium distributions far from the Boltzmann are often encountered. For example, power laws (kappa distributions) are common and could be used in place of the Boltzmann distribution [81]. There is no way of knowing in principle and practice guides what distribution is best used. Once the distribution is selected, the kinetic system will evolve and produce distributions far from the initial state.

Electrons and ions will not usually have the same moments. For two fluid models, the different electron and ion moments are known. In MHD, the local density n_s is assumed equal for both species (electrons and ions) to ensure quasi-neutrality. The local speed of both species is determined by taking together the local MHD speed (\mathbf{V}) and current density (\mathbf{J}):

$$\begin{aligned}\mathbf{v}_e &= \frac{\mathbf{J}m_e - \mathbf{V}q_i(m_en_e + m_in_i)}{m_en_e(q_e - q_i)} \\ \mathbf{v}_i &= \frac{-\mathbf{J}m_e + \mathbf{V}q_e(m_en_e + m_in_i)}{m_en_i(q_e - q_i)}\end{aligned}\quad (8.27)$$

where the mass, density, and charge of ions and electrons are used. The temperature is not defined completely. The MHD model uses a single scalar temperature. Occam's razor would suggest to assume equal temperature for both species, but in space typically, the barber is flamboyant and electrons and ions tend to have rather different temperatures. This can be accommodated by prescribing a rule for the temperature ratio derived from observations.

For more advanced fluid models with separate ion and electron temperature or even better with pressure tensors for both species, the information on the fluid moments is richer and can be more easily linked with kinetic models using sampling from the Chapman-Enskog expansion [82].

The initial state generated with this approach is not an exact solution of the Boltzmann equation because it only uses a limited number of moments. The Boltzmann distribution is a solution only if there are no spatial variations. The presence of spatial variations, even though self-consistent with the first moments, will lead to imbalances in the higher-order moments, for example, anisotropic components of the pressure tensor, heat flux, and higher. For this reason, after initialization, the kinetic simulation undergoes relaxation. Implicit and semi-implicit methods are advantageous in this situation because they tend to damp high-frequency and short-wavelength waves, as explained in Set. 8.1. This damping facilitates the elimination of noise and the establishment of a proper kinetic state after a short transient following the initial setup.

The most critical aspect of coupling fluid and kinetic models is the strategy for handling the interface conditions. Fluid models are structurally incapable of receiving the information from the kinetic models: for example, they can't handle waves that are not present in the fluid models or can't handle high-energy particle beams. We then use two types of boundary conditions, semi-open or linked with fluid. The semi-open boundary conditions are based on the perfectly matched layer [26, 83] approach where a damping term is added to the Maxwell's equations to absorb outgoing waves:

$$\frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{B} - \eta \mathbf{E} \quad (8.28)$$

A magnetic conductivity and a careful balance between that and the electric conductivity (or resistivity η) are needed to keep the refractive index equal to vacuum. The damping rate η increases monotonically from zero at the inner boundary of the damping region to its maximum value at the external boundary of the kinetic domain [83]. For the particles, the particles in the boundary layer are replaced at every time step with a new population based on the reference state provided by the MHD run. The thickness of the boundary layer can be varied

depending on the application. The MHD-driven boundary conditions are identical to the semi-open boundary conditions for the particles, but the fields are imposed from the MHD state.

8.5.2 One-Way or Two-Way Coupling

The MLMD concept has been applied to fluid-kinetic cooling both one way and two ways.

In the *one-way approach*, the kinetic state is spun off an evolving fluid state. As soon as started, the kinetic state will start to evolve differently from the fluid state. In one way coupling, there is no attempt to modify the fluid evolution to take into account the kinetic evolution. The complete details of the one-way coupling strategy are given in Ref. [84]. The idea of this approach is that kinetic models can only be run on small domains and most importantly for short intervals, making their contribution to the larger fluid simulation intermittent and local, fundamentally unsuitable to be fed back to the fluid macroscale, unless the so-called tooth-gap approach is used [85, 86] to nudge the fluid model toward the changes suggested by the fluid model. For this reason, considering the kinetic local evolution as a one-way spun-off simulation makes sense. The approach has been successfully applied to solar problems [87] and to magnetospheric problems [84] where the approach was used to couple one way the iPic3D code [15] with the UCLA global magnetospheric model [88, 89]. The approach successfully addresses the physics of substorms, reconnection, and particle energization in the Earth dayside [90] and nightside [91–93]. The approach has been validated with comparison with observational data [94, 95] and could supplement the global evolution with details of the electron scale physics [90, 96] observed by the MMS mission [97]. The approach is especially useful for handling particle acceleration and including all mechanisms of particle acceleration from macroscopic scales to the effect of wave-particle interaction at the microscopic level [90, 98].

In the *two-way coupling*, the results of the kinetic model are fed back to the fluid model. The complete details of the two-way coupling strategy are given in Ref. [99, 100]. This operation is in principle straightforward because the moments of the kinetic model directly give the needed information for the fluid model. But the kinetic level includes waves that are unknown to the fluid level, and even the known waves have different properties. The different response to electromagnetic fluctuations in fluid and kinetic models can result in fluid models not being able to properly evolve the information received from the kinetic level. An obvious example is that of the Hall electric and magnetic field produced by kinetic reconnection and absent from MHD models [99]: for this reason, coupling fluid and kinetic models is typically more successful if the fluid model being coupled is at least a Hall MHD description that can at least represent a whistler wave. But the same argument relative to Hall physics applies to other waves from Langmuir to Bernstein to anisotropy and multi-stream-driven modes. Fluid models completely miss that information. Another severe limitation of two-way coupling is that no high-energy

particle beam can be transferred to the fluid model. This can be especially damning if the beam in the larger fluid level would experience mirroring and return to the kinetic level (e.g., mirroring of electron jets in the polar regions of the Earth). A third limitation of two-way coupling is that for realistic parameters, the duration of a kinetic-level simulation is typically much shorter than for the fluid description. This is the main reason for preferring one-way coupling instead. But two-way coupling can still be valuable for problems where the jump between the sizes of the fluid and kinetic domains is modest or by using an unrealistic parameter that compress the scales between electrons and ions and between macroscopic and microscopic scales artificially [101]: this approach makes the two-way coupling viable. The approach has been used in practice to couple iPic3D [15] kinetic simulations with both Bats'R'us [102] and MPI-AMRVAC [100] with successful applications to planetary magnetospheres [103, 104] and key plasma processes [105, 106].

A different implementation of the fluid-kinetic coupling is based on using particle methods to describe also the fluid model. Particle in cell methods can be designed also to study fluid models; in this case, the particles represent Lagrangian elements of fluids [79, 107]. With this approach to fluid models, the linking of fluid and kinetic models becomes more natural as a particle can change its nature from kinetic (i.e., a Lagrangian element of phase space) to fluid (i.e., a Lagrangian fluid element) and vice versa depending on the location and grid resolution [108–110]. This approach is part of the larger class of *Asymptotic-Preserving* schemes where the transition from fluid to kinetic description is seamless and based on the resolution [111], applied with success also to plasma physics problems [112].

In all examples above, the fluid-kinetic coupling has benefited from the use of semi-implicit methods, in particular of the iPic3D code [15] based on the moment implicit method. The moment implicit method naturally damps high-frequency waves not resolved by the time step. In the examples above typically, the same time step is used in all levels of description, fluid and kinetic. The implicit and semi-implicit methods then eliminate the unresolved scales by applying numerical damping. This feature highly simplifies the coupling, but depending on the application, it might still be necessary to apply some filtering of the outgoing waves leaving the kinetic domain and entering the fluid domain.

Acknowledgments This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 776262 (AIDA, aida-space.eu) and No. 754304 (DEEP-EST, deep-projects.eu).

References

1. A.S.S. Lipatov, *The Hybrid Multiscale Simulation Technology* (Springer, Berlin, 2002)
2. G. Lapenta, in *Plasma Modeling: Methods and Applications*, ed. by G. Colonna, A. D'Angola (IOP Publishing, Bristol, 2016)
3. S. Markidis, G. Lapenta, J. Comput. Phys. **230**(18), 7037 (2011)
4. C. Birdsall, A. Langdon, *Plasma Physics via Computer Simulation* (Taylor & Francis, London, 2004)

5. R. Hockney, J. Eastwood, *Computer Simulation Using Particles* (Taylor & Francis, Milton Park, 1988)
6. Y.N. Grigoryev, V.A. Vshivkov, M.P. Fedoruk, *Numerical Particle-in-Cell Methods: Theory and Applications* (VSP BV, AH Zeist, 2005)
7. G. Lapenta, J. Comput. Phys. **231**(3), 795 (2012)
8. J.M. Hyman, M. Shashkov, J. Comput. Phys. **151**(2), 881 (1999)
9. J. Brackbill, D. Forslund, J. Comput. Phys. **46**, 271 (1982)
10. G. Lapenta, J. Comput. Phys. **334**, 349 (2017)
11. G. Chen, L. Chacón, D.C. Barnes, J. Comput. Phys. **230**(18), 7018 (2011)
12. G. Lapenta, D. Gonzalez-Herrero, E. Boella, J. Plasma Phys. **83**(2) (2017)
13. D. Gonzalez-Herrero, E. Boella, G. Lapenta, Comput. Phys. Commun. **229**, 162 (2018)
14. D. Gonzalez-Herrero, A. Micera, E. Boella, J. Park, G. Lapenta, Comput. Phys. Commun. **236**, 153 (2019)
15. S. Markidis, G. Lapenta, Rizwan-uddin, Math. Comput. Simul. **80**, 1509 (2010)
16. H.X. Vu, J.U. Brackbill, Comput. Phys. Comm. **69**, 253 (1992)
17. G. Lapenta, J. Brackbill, P. Ricci, Phys. Plasmas **13**(5), 055904 (2006)
18. A. Langdon, B. Cohen, A. Friedman, J. Comput. Phys. **51**, 107 (1983)
19. D.W. Hewett, A.B. Langdon, Comput. Phys. Commun. **48**(1), 127 (1988)
20. J.U. Brackbill, B.I. Cohen (eds.), *Multiple Time Scales* (Academic Press, Cambridge, 1985)
21. J.U. Brackbill, J. Comput. Phys. **317**, 405 (2016)
22. J. Villasenor, O. Buneman, Comput. Phys. Commun. **69**(2–3), 306 (1992)
23. T. Umeda, Y. Omura, T. Tominaga, H. Matsumoto, Comput. Phys. Commun. **156**(1), 73 (2003)
24. T.Z. Esirkepov, Comput. Phys. Commun. **135**(2), 144 (2001)
25. G. Chen, L. Chacon, Comput. Phys. Commun. **197**, 73 (2015)
26. A. Taflove, S.C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method* (Artech House Publishers, Norwood, 2005)
27. R.L. Morse, C.W. Nielson, Phys. Fluids **14**, 830 (1971). <https://doi.org/10.1063/1.1693518>
28. E. Lifschitz, L. Pitajewski, in *Textbook of Theoretical Physics*, vol. 10 (Pergamon Press, Oxford, 1983)
29. D.G. Swanson, *Plasma Waves*, 2nd ed. Plasma Physics (IOP, Bristol, 2003)
30. J. Brackbill, G. Lapenta, J. Comput. Phys. **114**(1), 77 (1994)
31. J.L. Vay, P. Colella, J.W. Kwan, P. McCorquodale, D.B. Serafini, A. Friedman, D.P. Grote, G. Westenskow, J.C. Adam, A. Héron, I. Haber, Phys. Plasmas **11**, 2928 (2004)
32. M.J. Berger, P. Colella, J. Comput. Phys. **82**(1), 64 (1989)
33. G. Jacobs, J.S. Hesthaven, J. Comput. Phys. **214**(1), 96 (2006)
34. V.D. Liseikin, *Grid Generation Methods* (Springer, Berlin, 1999)
35. M. Ainsworth, J.T. Oden, *A Posteriori Error Estimation in Finite Element Analysis* (Wiley, Hoboken, 2000)
36. G. Lapenta, Int. J. Num. Meth. Eng. **59**, 2065 (2004)
37. J. Brackbill, J. Comput. Phys. **108**(1), 38 (1993)
38. G. Lapenta, J. Comput. Phys. **193**(1), 159 (2004)
39. J.U. Brackbill, J.S. Saltzman, J. Comput. Phys. **46**(3), 342 (1982)
40. G.L. Delzanno, L. Chacón, J.M. Finn, Y. Chung, G. Lapenta, J. Comput. Phys. **227**(23), 9841 (2008)
41. D. Sulsky, J.U. Brackbill, J. Comput. Phys. **96**, 339 (1991)
42. G. Lapenta, J. Comput. Phys. **230**(12), 4679 (2011)
43. L. Chacón, G. Chen, J. Comput. Phys. **316**, 578 (2016)
44. M. Desbrun, A.N. Hirani, M. Leok, J.E. Marsden, preprint, arXiv:math/0508341 (2005)
45. J.L. Vay, P. Colella, A. Friedman, D.P. Grote, P. McCorquodale, D. Serafini, Comput. Phys. Commun. **164**(1–3), 297 (2004)
46. A. Allievi, R. Bermejo, J. Comput. Phys. **132**(2), 157 (1997)
47. G.L. Delzanno, E. Camporeale, J.D. Moulton, J.E. Borovsky, E.A. MacDonald, M.F. Thomson, IEEE Trans. Plasma Sci. **41**(12), 3577 (2013)

48. B. van der Holst, R. Keppens, *J. Comput. Phys.* **226**(1), 925 (2007)
49. A. Mignone, C. Zanni, P. Tzeferacos, B. Van Straalen, P. Colella, G. Bodo, *Astrophys. J. Suppl. Ser.* **198**(1), 7 (2011)
50. G. Tóth, B. Van der Holst, I.V. Sokolov, D.L. De Zeeuw, T.I. Gombosi, F. Fang, W.B. Manchester, X. Meng, D. Najib, K.G. Powell, et al., *J. Comput. Phys.* **231**(3), 870 (2012)
51. J. Vay, P. Colella, P. McCrorquodale, B. van Straalen, A. Friedman, D.P. Grote, *Laser Particle Beams* **20**, 569 (2002). <https://doi.org/10.1017/S0263034602204139>
52. P. Colella, P.C. Norgaard, *J. Comput. Phys.* **229**, 947 (2010). <https://doi.org/10.1016/j.jcpc.2009.07.004>
53. K. Fujimoto, R.D. Sydora, *Comput. Phys. Commun.* **178**, 915 (2008). <https://doi.org/10.1016/j.cpc.2008.02.010>
54. T. Matsui, H. Usui, T. Moritaka, M. Nunami, in *2011 19th International Euromicro Conference on Parallel, Distributed and Network-Based Processing* (IEEE, Piscataway, 2011), pp. 277–281
55. A. Beck, M.E. Innocenti, G. Lapenta, S. Markidis, *J. Comput. Phys.* **271**, 430 (2014)
56. M.E. Innocenti, G. Lapenta, S. Markidis, A. Beck, A. Vapirev, *J. Comput. Phys.* **238**, 115 (2013)
57. M.E. Innocenti, A. Johnson, S. Markidis, J. Amaya, J. Deca, V. Olshevsky, G. Lapenta, *Adv. Eng. Softw.* **111**, 3 (2017)
58. M.E. Innocenti, A. Tenerani, M. Velli, *Astrophys. J.* **870**(2), 66 (2019)
59. M.E. Innocenti, A. Beck, S. Markidis, G. Lapenta, *J. Comput. Phys.* **312**, 14 (2016)
60. M. Innocenti, C. Tronci, S. Markidis, G. Lapenta, in *Journal of Physics: Conference Series*, vol. 719 (IOP Publishing, Bristol, 2016), p. 012019
61. M.E. Innocenti, A. Beck, T. Ponweiser, S. Markidis, G. Lapenta, *Comput. Phys. Commun.* **189**, 47 (2015)
62. M. Innocenti, C. Norgren, D. Newman, M. Goldman, S. Markidis, G. Lapenta, *Phys. Plasmas* **23**(5), 052902 (2016)
63. G. Lapenta, J.U. Brackbill, *J. Comput. Phys.* **115**, 213 (1994)
64. D.P. Landau, K. Binder, *Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2000)
65. G. Lapenta, *J. Comput. Phys.* **181**(1), 317 (2002)
66. G. Lapenta, J. Brackbill, *Comput. Phys. Commun.* **87**(1–2), 139 (1995)
67. J. Frederiksen, G. Lapenta, M. Pessah, preprint, arXiv:1504.03849 (2015)
68. P.T. Luu, T. Tückmantel, A. Pukhov, *Comput. Phys. Commun.* **202**, 165 (2016)
69. M. Drouin, preprint, arXiv:1805.06540 (2018)
70. D.R. Welch, T.C. Genoni, R.E. Clark, D.V. Rose, *J. Comput. Phys.* **227**(1), 143 (2007)
71. T. Grismayer, M. Vranic, J.L. Martins, R. Fonseca, L.O. Silva, *Phys. Rev. E* **95**(2), 023210 (2017)
72. C. Li, J. Teunissen, M. Nool, W. Hundsdorfer, U. Ebert, *Plasma Sources Sci. Technol.* **21**(5), 055019 (2012)
73. J. Teunissen, U. Ebert, *J. Comput. Phys.* **259**, 318 (2014)
74. F. Assous, T.P. Dulimbart, J. Segré, *J. Comput. Phys.* **187**(2), 550 (2003)
75. M. Vranic, T. Grismayer, J.L. Martins, R.A. Fonseca, L.O. Silva, *Comput. Phys. Commun.* **191**, 65 (2015)
76. K. Huang, *Introduction to Statistical Physics* (Chapman and Hall/CRC, London, 2009)
77. J. Burgers, *Flow Equations for Composite Gases* (Academic Press, New York, 1969)
78. H. Alfvén, *Nature* **150**(3805), 405 (1942)
79. V. Olshevsky, F. Bacchini, S. Poedts, G. Lapenta, *Comput. Phys. Commun.* **235**, 16 (2019)
80. C. Cercignani, in *The Boltzmann Equation and Its Applications* (Springer, Berlin, 1988), pp. 40–103
81. V. Pierrard, M. Lazar, *Solar Phys.* **267**(1), 153 (2010)
82. A.L. Garcia, B.J. Alder, *J. Comput. Phys.* **140**(1), 66 (1998)
83. J.P. Berenger, *J. Comput. Phys.* **114**(2), 185 (1994)
84. R.J. Walker, G. Lapenta, J. Berchem, M. El-Alaoui, D. Schriver, *J. Plasma Phys.* **85**(1) (2019)

85. I.G. Kevrekidis, C.W. Gear, J.M. Hyman, P.G. Kevrekidis, O. Runborg, C. Theodoropoulos, et al., *Commun. Math. Sci.* **1**(4), 715 (2003)
86. G. Samaey, D. Roose, I.G. Kevrekidis, *Multiscale Model. Simul.* **4**(1), 278 (2005)
87. G. Baumann, T. Haugbølle, Å. Nordlund, *Astrophys. J.* **771**(2), 93 (2013)
88. J. Raeder, R. Walker, M. Ashour-Abdalla, *Geophys. Res. Lett.* **22**(4), 349 (1995)
89. M. El-Alaoui, *J. Geophys. Res. Space Phys.* **106**(A4), 6229 (2001)
90. G. Lapenta, J. Berchem, M. Zhou, R. Walker, M. El-Alaoui, M. Goldstein, W. Paterson, B. Giles, C. Pollock, C. Russell, et al., *J. Geophys. Res. Space Phys.* **122**(2), 2024 (2017)
91. M. Ashour-Abdalla, G. Lapenta, R.J. Walker, M. El-Alaoui, H. Liang, *J. Geophys. Res. Space Phys.* **120**(6), 4784 (2015)
92. M. Ashour-Abdalla, G. Lapenta, R. Walker, M. El-Alaoui, H. Liang, M. Zhou, J. Berchem, M.L. Goldstein, *Geophys. Res. Lett.* **43**(12), 6005 (2016)
93. R.J. Walker, G. Lapenta, H. Liang, J. Berchem, M. El-Alaoui, M.L. Goldstein, *J. Geophys. Res. Space Phys.* **123**(10), 8241 (2018)
94. M. Zhou, J. Berchem, R. Walker, M. El-Alaoui, X. Deng, E. Cazzola, G. Lapenta, M. Goldstein, W. Paterson, Y. Pang, et al., *Phys. Rev. Lett.* **119**(5), 055101 (2017)
95. M. Zhou, J. Berchem, R. Walker, M. El-Alaoui, M. Goldstein, G. Lapenta, X. Deng, J. Li, O. Le Contel, D. Graham, et al., *J. Geophys. Res. Space Phys.* **123**(3), 1834 (2018)
96. M. Zastrow, *Eos* **98** (2017). <https://doi.org/10.1029/2017EO072923>
97. J. Burch, R. Torbert, T. Phan, L.J. Chen, T. Moore, R. Ergun, J. Eastwood, D. Gershman, P. Cassak, M. Argall, et al., *Science* **352**(6290), aaf2939 (2016)
98. G. Lapenta, M. Ashour-Abdalla, R.J. Walker, M. El Alaoui, *Geophys. Res. Lett.* **43**(2), 515 (2016)
99. L.K. Daldorff, G. Tóth, T.I. Gombosi, G. Lapenta, J. Amaya, S. Markidis, J.U. Brackbill, *J. Comput. Phys.* **268**, 236 (2014)
100. K. Makwana, R. Keppens, G. Lapenta, *Comput. Phys. Commun.* **221**, 81 (2017)
101. G. Tóth, Y. Chen, T.I. Gombosi, P. Cassak, S. Markidis, I.B. Peng, *J. Geophys. Res. Space Phys.* **122**(10) (2017)
102. G. Tóth, I.V. Sokolov, T.I. Gombosi, D.R. Chesney, C.R. Clauer, D.L. De Zeeuw, K.C. Hansen, K.J. Kane, W.B. Manchester, R.C. Oehmke, K.G. Powell, A.J. Ridley, I.I. Roussev, Q.F. Stout, O. Volberg, R.A. Wolf, S. Sazykin, A. Chan, B. Yu, J. Kóta, *J. Geophys. Res. Space Phys.* **110**(A9), 12226 (2005). <https://doi.org/10.1029/2005JA011126>
103. G. Tóth, X. Jia, S. Markidis, I.B. Peng, Y. Chen, L.K. Daldorff, V.M. Tenishev, D. Borovikov, J.D. Haiducek, T.I. Gombosi, et al., *J. Geophys. Res. Space Phys.* **121**(2), 1273 (2016)
104. Y. Chen, G. Tóth, P. Cassak, X. Jia, T.I. Gombosi, J.A. Slavin, S. Markidis, I.B. Peng, V.K. Jordanova, M.G. Henderson, *J. Geophys. Res. Space Phys.* **122**(10), 10 (2017)
105. K. Makwana, R. Keppens, G. Lapenta, in *Journal of Physics: Conference Series*, vol. 1031 (IOP Publishing, Bristol, 2018), p. 012019
106. K. Makwana, R. Keppens, G. Lapenta, *Phys. Plasmas* **25**(8), 082904 (2018)
107. J. Brackbill, *J. Comput. Phys.* **96**(1), 163 (1991)
108. S. Markidis, P. Henri, G. Lapenta, K. Rönnmark, M. Hamrin, Z. Meliani, E. Laure, *J. Comput. Phys.* **271**, 415 (2014)
109. S. Markidis, V. Olshevsky, C.P. Sishtla, S.W.d. Chien, G. Lapenta, E. Laure, *Front. Physiol.* **6**, 100 (2018)
110. N. Horsten, G. Samaey, M. Baelmans, *Nucl. Mat. Energ.* **18**, 201 (2019)
111. S. Jin, *SIAM J. Sci. Comput.* **21**(2), 441 (1999)
112. P. Degond, F. Deluzet, *J. Comput. Phys.* **336**, 429 (2017)



Hybrid-Kinetic Approach: Inertial Electrons

9

Neeraj Jain, Patricio A. Muñoz, and Jörg Büchner

Abstract

Hybrid-kinetic simulations describe ion-scale kinetic phenomena in space plasmas by considering ions kinetically, i.e., as particles, while electrons are modeled as a fluid. Most of the existing hybrid-kinetic codes neglect the electron mass (see Chap. 3) for a simplified calculation of the electromagnetic fields. There are, however, situations in which delay in the electrons response due to the electron inertia matters. This chapter concentrates on hybrid-kinetic simulation models that take the finite mass of the electron fluid into account. First, a review is given of the history of including the finite electron mass in hybrid-kinetic models. Then the equations are discussed which additionally have to be solved compared to the massless hybrid-kinetic models. For definiteness, their numerical implementation without additional approximations is illustrated by describing a hybrid-kinetic code, CHIEF. The importance of the consideration of the finite electron mass are discussed for typical applications (magnetic reconnection, plasma turbulence, collisionless shocks, and global magnetospheric simulations). In particular, the problem of guide field magnetic reconnection is addressed in some detail. Possible next steps toward further improvements of hybrid-kinetic simulations with finite electron mass are suggested.

N. Jain (✉)

Center for Astronomy and Astrophysics, Technical University Berlin, Berlin, Germany
e-mail: n.jain@astro.physik.tu-berlin.de

P. A. Muñoz · J. Büchner

Max-Planck-Institute for Solar System Research, Göttingen, Germany
e-mail: munozp@mps.mpg.de; buechner@tu-berlin.de; buechner@mps.mpg.de

9.1 Introduction

The relative scarcity of binary particle collisions in space and astrophysical plasmas makes many of their plasma processes (like magnetic reconnection, shock waves, and turbulence) multi-scale in nature, spanning from the largest, global fluid to electron kinetic scales. In order to investigate those usually nonlinear and nonlocal processes by computer simulations, one needs to resolve, in principle, the smallest scales necessary for the processes and, at the same time, the large scales of the system, both in space and time. While fully kinetic plasma models well describe physical processes down to ion and electron kinetic scales, they require prohibitively expensive computational resources to describe at the same time larger-scale processes even for future computational facilities.

The physicist's way out of this situation is to understand the processes in parts by modelling and simulating certain aspects of physics while neglecting others. This implies the use of a variety of simulation models. A comprehensive understanding is then obtained by combining various aspects of these simulations.

In the first part of the book, the basic plasma approximations and models for the simulation of space and astrophysical plasmas were discussed: magnetohydrodynamic (MHD, Chap. 1), Hall-MHD (Chap. 2), hybrid-kinetic with massless electrons (Chap. 3), gyro-kinetic (Chap. 4), Eulerian Vlasov (Chap. 5), and Particle in Cell (PIC, Chap. 6) simulation models. These basic simulation models and codes separately cover all different ranges of scales and physics. Kinetic scale processes taking place at ion and electron spatiotemporal scales are of critical importance for the mostly collisionless space and astrophysical plasmas. For example, collisionless magnetic reconnection, considered to be responsible for explosive energy release in many space and astrophysical environments including large-scale substorms in the Earth's magnetosphere as well as solar and stellar flares, is enabled by essentially kinetic scale processes [1]. The solar wind turbulence (magnetic field fluctuations) is dissipated by kinetic scales processes [2], and collisionless shock waves essentially require kinetic scale dissipation processes [3]. Simulations of kinetic scale processes are, therefore, indispensable to understand the physics of space and astrophysical plasmas.

Eulerian Vlasov and semi-Lagrangian PIC codes (Chaps. 5 and 6), in principle, fully describe ion- and electron-scale kinetic plasma processes. The computational feasibility, however, restricts their use to artificial physical parameters, which reduce, e.g., the gap between electron and ion scales [4]. Fully kinetic simulations are, therefore, usually carried out for ion-to-electron mass ratios in the range 25–400 while for a full-scale separation between ion- and electron kinetic processes, mass ratios of the order of 1800 would be necessary.

Furthermore, PIC code simulations (Chap. 6) suffer from their numerical shot noise due to a finite number of computationally traced particles compared to the much larger number of particles in real space and astrophysical plasma systems [5, 6]. Vlasov code simulations (Chap. 5) are free from such particle-associated noise, but they, on the other hand, are computationally equally challenging as PIC code

simulations due to the necessity to resolve particle distribution functions on a fine grid in the velocity space [7, 8].

Gyro-kinetic plasma models (Chap. 4) reduce the dimensions of the velocity space by one, making simulations this way computationally less demanding. But they are restricted, e.g., by not resolving the details of the electron and ion temporal processes [9] and being valid in strongly magnetized plasmas like in magnetically confining fusion devices and, maybe, in pulsars and accretion disks.

Hybrid-kinetic plasma simulation models completely neglect kinetic effects associated with the electrons. This further reduces the computational burden (see, e.g., [10]). Such models are, therefore, suitable to study processes for which ion kinetic effects are important while electron kinetic effects are not.

The majority of existing hybrid-kinetic plasma simulation models simplify their numerical implementation by assuming a massless (inertia-less) electron fluid. This assumption allows a direct calculation of the electric field from the electron's momentum equation without the need of solving partial differential equations for the electric field (for details, see Chap. 3). It, however, limits the validity of the model to processes at scales exceeding by far the electron scales.

Hybrid-kinetic codes with massless electrons have been used to simulate global phenomena like the formation of foreshocks and cavitons by the interaction of the solar wind with the Earth's magnetosphere [11, 12], day side magnetic reconnection through the Earth's magnetopause [13, 14], the formation of filamentary structures in the Earth's magnetosheath [15], and astrophysical explosions [16] as well as for specific physics studies of magnetic reconnection [17–19], plasma turbulence [20, 21], and shock waves [22]). They contributed to the understanding of space and astrophysical plasma phenomena at scales of the order of the ion scale or larger.

Effects of the finite electron mass originate from electron scales. They nevertheless may play an important role for larger-scale plasma phenomena as well. Laboratory experiments of reconnection have revealed that the thickness of current sheets may be as thin as several electron inertial lengths [23]. At the same time, the current sheet thickness can be larger than the electron gyro-radius. This allows a fluid description of the electrons. And hybrid-kinetic simulations with a finite-mass electron fluid have shown that the electron inertia can dominate when the current sheets thin down to electron inertial scale lengths [17, 18]. In hybrid-kinetic simulations of collisionless plasma turbulence, carried out for massless electrons, current sheets thin down to the grid scale [21, 24]. This is not physical. Hence, hybrid-kinetic simulation studies of thinning current sheets must incorporate electron inertial effects. Hybrid-kinetic simulations with inertial electrons indeed show the formation of electron-scale thin current sheets in collisionless plasma turbulence. Shock steepening to the grid scale has also been observed in hybrid-kinetic simulations (with massless electrons) of nonstationary shocks [25]. Inertial electron hybrid-kinetic simulations showed that the finite mass of the electrons determines the macroscopic shock reformation [26]. In particular, it determines the strong phase-space mixing between the upstream (incoming) and the reflected ions via wave-particle interactions, the ion thermalization, and the growth of plasma waves in the foot and ramp regions of the shock.

Thus, the consideration of the finite electron mass, although it is small, is important to understand a number of critical phenomena in space and astrophysical plasmas.

This chapter describes hybrid-kinetic plasma models that take into account the finite mass of the electrons. First, in Sect. 9.2, a review of the history of the development of finite-mass electron codes is given. In Sect. 9.3, we present the equations which have to be solved. As an example, we discuss in Sect. 9.4 their numerical solution by the code CHIEF [27], their numerical implementation, the parallelization of the code, and its performance and scaling. In Sect. 9.5, the importance of the consideration of the finite electron mass in hybrid simulations are discussed, comparing results of hybrid code reconnection simulations with stationary and nonstationary ions. We also discuss other applications like collisionless shocks, plasma turbulence, and global magnetospheric simulations. Future improvements of related algorithms are discussed in Sect. 9.6.

9.2 Historical Development of Finite-Electron-Mass Hybrid-Kinetic Simulation Models

A number of hybrid-kinetic models have been developed that include different terms related to the electron inertia under varying degrees of approximations.

First hybrid-PIC models with ions described as macro-particles (PIC method) and electrons modeled as a fluid were developed together with the development of fully kinetic PIC codes [28, 29]. Those first hybrid-PIC codes used the Darwin approximation and solved the electromagnetic field equations separately for the curl-free and the divergence-free parts of the potentials as well as for the current density. In this approximation, the electron inertia had to be considered only for calculating the transverse electron current density. Practically, hybrid codes based on the Darwin approximation have rarely been used, except for 1D cases in which they calculated only electrostatic potentials.

Then hybrid code algorithms with electron inertia were developed to globally simulate the entire Earth's magnetosphere [30, 31]. In this approach, the electric field is derived from the electron momentum equation, the electron velocity is obtained by solving Ampère's law, and the magnetic field is obtained by solving Faraday's law. The electron inertia itself was considered only as a correction by taking into account the electron polarization drift in the (implicit) equation for the magnetic field while it was neglected in all other equations.

Later hybrid codes with electron inertia solved for the generalized electromagnetic fields that satisfy Faraday's law like equations (see Sec. 5.7 in [10]; see also [17, 32]). These models obtained solutions of the generalized electric and magnetic field equations by means of predictor-corrector schemes, which require a staggered grid, or by a trapezoidal leapfrog algorithm. The electromagnetic fields were then obtained from the generalized fields under different approximations. The magnetic field is calculated solving an elliptic partial differential equation (PDE) obtained from the expression for the generalized magnetic field neglecting electron-

scale density variations. The electric field was then calculated from the generalized Ohm's law by neglecting the electron inertial term with time derivatives of the electron fluid velocity [17]. Other authors neglected even the convective electron acceleration term [32].

These hybrid codes that partially included electron inertial effects have mainly been used to study collisionless magnetic reconnection [33–35]. In particular, Shay et al. (1998) used an evolution equation for a scalar electron pressure [32], while Kuznetsova et al. (1998) included the full electron pressure tensor to take into account the non-gyrotropic effects [17]. The latter were shown to play an important role in possibly balancing the reconnection electric field.

Another method of calculating the electric field is the solution of an elliptic PDE for the electric field, obtained by combining the Ohm's law with Maxwell's equations (see Sec. 5.2.4 of [10]). This approach was utilized for a one-dimensional finite-electron-mass hybrid code [36]. In that code, electric and magnetic fields were obtained by solving the elliptic PDE and Faraday's law, respectively, and not by calculating for the generalized electromagnetic fields as in Refs. [10, 17, 32]. Electron inertia effects were considered in the elliptic equation for the electric field, while, still, the electron inertia term was ignored that contains the divergence of the electric field. This code used a variable mass ratio in order to better model low-density plasma regions [36]. The mass ratio variation was implemented in the code by locally and temporally varying the electron mass, adjusting it to always satisfy the CFL condition based on the electron Alfvén speed for a given time step (and grid cell size). This works well as long as the scales of interest are not too close to the electron inertial scales.

More recently, a hybrid-PIC model was developed that implemented electron inertial terms without any of the approximations used in the codes before (see above) [27]. In Sec. 9.4, the detailed algorithm and implementation of this code CHIEF (Code Hybrid with Inertial Electron Fluid) will be described. Some applications which illustrate the need of considering the full electron inertia for describing guide field magnetic reconnection will be discussed in Sect. 9.5.1.

Hybrid-Vlasov codes directly obtain the ion velocity space distributions by an Eulerian solution of the Vlasov equations, while electrons are described as an inertial fluid (see, e.g., [37]). In this approach, the ion Vlasov equations are solved together with the electron fluid and Maxwell's equations for the electromagnetic fields. For the code described by Valentini et al. (2007), a Helmholtz equation for the electric field was obtained by taking the curl of Faraday's law, using a generalized Ohm's law and ignoring density perturbations contained in one of the electron inertial terms [37]. This code has been applied, e.g., to describe the solar wind turbulence [38–40] as well as to other 2D3V and 3D3V plasma problems [41–43].

Cheng et al. (2013) developed a three-dimensional hybrid-kinetic code with electron inertia utilizing for the ions a δf method, i.e., evolving only the variations of the ion distribution function that have to be assumed to be small compared to a given background [44]. In this δf code, the electron inertia was included in the generalized Ohm's law for the electric field, calculated using the Ampère's

law and the ion fluid momentum equation. Such approach was first proposed by Jones et al. in 2003 for solving the 1D gyrofluid electron equations [45]. It was applied to study the propagation of dispersive Alfvén waves in the coupled Earth's magnetosphere-ionosphere system and electron acceleration [46, 47] as well as to describe the Alfvén wave dynamics in the Io-Jupiter flux tube [48].

9.3 Equations to be Solved

Here, we describe hybrid-PIC models in which ions are treated as particles and electrons as a fluid. The equations of motion of every ion macro-particle,

$$\frac{d\mathbf{x}_p}{dt} = \mathbf{v}_p \quad \text{and} \quad (9.1)$$

$$m_i \frac{d\mathbf{v}_p}{dt} = e(\mathbf{E} + \mathbf{v}_p \times \mathbf{B}), \quad (9.2)$$

where \mathbf{x}_p and \mathbf{v}_p are the positions and velocities of ion macro-particles, respectively, and are solved by a PIC algorithm (see Chap. 6).

The full finite-mass electron fluid momentum equation is given by

$$m_e \left[\frac{\partial \mathbf{u}_e}{\partial t} + (\mathbf{u}_e \cdot \nabla) \mathbf{u}_e \right] = -e(\mathbf{E} + \mathbf{u}_e \times \mathbf{B}) - \frac{1}{n_e} \nabla \cdot \underline{\mathbf{P}}_e, \quad (9.3)$$

where \mathbf{u}_e and $\underline{\mathbf{P}}_e$ are the electron fluid velocity and the electron pressure tensor, respectively. The electron density is determined by the quasi-neutrality condition $n_e = n_i = n$. The electric and magnetic fields, \mathbf{E} and \mathbf{B} , are related to the plasma current density via Maxwell's equations:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (9.4)$$

$$\nabla \times \mathbf{B} = \mu_0 n e (\mathbf{u}_i - \mathbf{u}_e) \quad (9.5)$$

Here, \mathbf{u}_i is the ion fluid velocity obtained by the first-order velocity moment of the ion distribution function. From Eqs. 9.3 and 9.4, we obtain an evolution equation for the generalized vorticity $\mathbf{W} = \nabla \times \mathbf{u}_e - e\mathbf{B}/m_e$.

$$\frac{\partial \mathbf{W}}{\partial t} = \nabla \times (\mathbf{u}_e \times \mathbf{W}) - \nabla \times \left(\frac{\nabla p_e}{m_e n_e} \right) \quad (9.6)$$

$$p_e = n_e T_e \quad (9.7)$$

For simplicity, usually the electron pressure is assumed to be a scalar quantity p_e . For a uniform temperature T_e , the last term in Eq. 9.6 vanishes.

Substituting \mathbf{u}_e in the expression of \mathbf{W} from Ampere's law, one obtains an elliptic equation for the magnetic field,

$$\frac{1}{\mu_0 e} \nabla \times \left(\frac{\nabla \times \mathbf{B}}{n} \right) + \frac{e \mathbf{B}}{m_e} = \nabla \times \mathbf{u}_i - \mathbf{W} \quad (9.8)$$

Equations 9.1–9.8 represent a complete hybrid-PIC plasma model without any approximation on the electron inertial terms in the electron momentum equation. Note that in conventional hybrid codes, which neglect the electron inertia, the terms on the left hand side (LHS) of Eq. 9.3 are neglected, allowing a direct calculation of \mathbf{E} from this equation. As a result, the $\nabla \times \mathbf{u}_e$ term in the expression of \mathbf{W} disappears and Eq. (9.6) becomes an evolution equation for the magnetic field, i.e., Eq. (9.8) does not have to be solved at all.

In conventional hybrid codes with electron inertia, the electric and magnetic fields are obtained from Eqs. 9.3 and 9.6 by usually neglecting some of the terms. Substituting for \mathbf{u}_e from Eq. 9.5 and neglecting terms proportional to $\partial \mathbf{u}_i / \partial t$ and $\partial n / \partial t$, the LHS of Eq. 9.6 can be written entirely in terms of the time derivative of the magnetic field. The neglect of $\partial \mathbf{u}_i / \partial t$ and $\partial n / \partial t$ in Eq. 9.6 is justified for a large mass ratio m_i / m_e . However, these approximations are not necessarily valid in all situations of interest and need to be separately justified, especially for artificially small mass ratios m_i / m_e . Neglecting electron-scale variation of the density, the LHS of that equation can further be simplified to $\partial / \partial t [e(\mathbf{B} - d_e^2 \nabla^2 \mathbf{B}) / m_e]$, where $d_e = c / \omega_{pe}$ is the electron inertial length. The electric field is calculated directly from Eq. 9.3 by neglecting $\partial \mathbf{u}_e / \partial t$, which is clearly inconsistent with keeping this term to obtain Eq. 9.6.

In another approach, the electric field is calculated from Eq. 9.3 by explicitly evaluating the time derivative term $\partial \mathbf{u}_e / \partial t$ [27]. For this purpose, Munoz et al. in 2018 used the heaviness of the ions to assume that the ion quantities do not change much during a single time step, i.e., during a fraction of the electron gyro-period [27]. Thus, for an update of ion positions and velocities by a single time step dt , \mathbf{u}_e at $t = t_0 + dt$ and $t = t_0 + 2dt$ can be obtained by advancing Eq. 9.6, first from $t = t_0$ to $t = t_0 + dt$ and then from $t = t_0 + dt$ to $t = t_0 + 2dt$. This allows the calculation of $\partial \mathbf{u}_e / \partial t$ by a central difference scheme. In the second of the two time steps, the ion current and density from the previous time step can be used. Alternatively, the electric field can be calculated by solving an elliptical PDE, which one obtains by taking the curl of the Faraday's law (Eq. 9.4), using Ampère's law (Eq. 9.5) and substituting for $\partial \mathbf{J} / \partial t$ from the generalized Ohm's law. The numerical solution of the resulting elliptical PDE in more than one dimension is, however, much more involved than that of the elliptic PDE for the magnetic field (Eq. 9.8). The difficulty arises from the fact that one cannot use a Poisson equation to substitute for $\nabla \cdot \mathbf{E}$, which yields cross-derivative terms in the equation. In case of Eq. 9.8, the condition $\nabla \cdot \mathbf{B} = 0$ simplifies the calculations.

9.4 Numerical Implementation

For definiteness, we illustrate how a hybrid-PIC simulation model with a finite-mass electron fluid without additional approximations (see Sec. 9.3) works, by describing the model's numerical implementation in a code called CHIEF (Code Hybrid with Inertial Electron Fluid). CHIEF combines elements of the PIC code ACRONYM to describe the ions as particles (see, e.g., [49]¹) with an EMHD code that solves the electromagnetic equations coupled with an inertial electron fluid (see, e.g., [50]). Both codes have separately been tested and independently used to simulate different aspects of magnetic reconnection [51, 52], plasma instabilities [53, 54], particle acceleration [55], shocks [56], wave coupling [57], resonant wave-particle interactions [58], etc. These simulations have helped to clean up both the codes off numerical errors.

9.4.1 Ions as Particles

In the ion-related (PIC-) part of the code, the macro-particle positions and velocities at the N th time step are advanced for the electric and magnetic field values known at the time step N and defined on a staggered grid, the Yee lattice grid [59]. For this, the fields are first interpolated from the grid to the (macro-) ions positions with a weighting given by a shape function $S(\mathbf{x} - \mathbf{x}^p)$:

$$\mathbf{E}^p = \mathbf{E}(\mathbf{x}^p) = \int \mathbf{E}(\mathbf{x}) S(\mathbf{x} - \mathbf{x}^p) d\mathbf{x}, \quad (9.9)$$

$$\mathbf{B}^p = \mathbf{B}(\mathbf{x}^p) = \int \mathbf{B}(\mathbf{x}) S(\mathbf{x} - \mathbf{x}^p) d\mathbf{x}, \quad (9.10)$$

where the superscript p indicates the location of each (macro-) particle. Then, the ions are moved by a second-order accurate leapfrog algorithm. This means that at each time step N , the ion velocities are advanced from the half time step $N - 1/2$ to the half time step $N + 1/2$ and the ion positions from time step N to time step $N + 1$ by using the discretized version of Eqs. 9.1 and 9.2 (see, e.g., Sec. 4.3 of [60]):

$$\mathbf{x}^{p,N+1} = \mathbf{x}^{p,N} + \mathbf{v}^{p,N+1/2} \Delta t, \quad (9.11)$$

$$\mathbf{v}^{p,N+1/2} = \mathbf{v}^{p,N-1/2} + \frac{e \Delta t}{m_i} \left(\mathbf{E}^{p,N} + \frac{\mathbf{v}^{p,N+1/2} + \mathbf{v}^{p,N-1/2}}{2} \times \mathbf{B}^{p,N} \right) \quad (9.12)$$

Note that the first equation is explicit for $\mathbf{x}^{p,N+1}$ while the second is implicit for $\mathbf{v}^{p,N+1/2}$. The ion velocity is advanced via a Boris method [61], i.e., by a rotation

¹ <http://plasma.nerd2nerd.org/>.

of $\mathbf{v}^{p,N+1/2}$ (i.e., the magnitude of the vector stays constant), making Eq. 9.12 also explicit.

Because of the staggered position and velocity updates, the sources of the electromagnetic field, the ion number density n_i and ion current density \mathbf{j}_i , are computed at time steps indexed $N + 1$ and $N + 1/2$, respectively. The deposition of both the ion number and current density onto the grid is done via an interpolation scheme using the same shape function as for the electromagnetic field interpolation from the grid to the macro-particles position (Eqs. 9.9–9.10):

$$n_i^{N+1}(\mathbf{x}) = \sum_p N_p S(\mathbf{x} - \mathbf{x}^{p,N+1}), \quad (9.13)$$

$$\mathbf{j}_i^{N+1/2}(\mathbf{x}) = \sum_p N_p \mathbf{v}^{p,N+1/2} S(\mathbf{x} - \mathbf{x}^{p,N+1}). \quad (9.14)$$

The use of the same shape function in Eqs. 9.9 and 9.13 avoids self-forces and preserves the global momentum (see, e.g., Secs. 8.5–6 in [60] and Sec 5.3.3 of [62] and [63] for fully kinetic PIC codes or Sec. 4.5.2 in [10] for hybrid-PIC codes).

9.4.2 Electron Fluid

The electron density $n_e^{N+1/2}$ is set equal to the ion density $n_i^{N+1/2} = (n_i^N + n_i^{N+1})/2$, i.e., $n_e^{N+1/2}$ is obtained by applying the quasi-neutrality condition. If a constant (not evolving in time) electron temperature T_e is assumed, the electron pressure $p_e^{N+1/2} = n_e^{N+1/2} k_B T_e$ can be calculated via the equation of state, Eq. 9.7.

9.4.3 Electromagnetic Fields

Once the sources of the electromagnetic fields (Eqs. 9.13–9.14) are known on the Yee lattice, the electromagnetic fields are updated by using the electron fluid equations and Maxwell's equations. The electric and magnetic fields can be updated, e.g., by advancing the generalized vorticity \mathbf{W} in Eq. 9.6 from time step N to $N + 1$ using a flux-corrected transport algorithm like the one provided by the LCPFCT package.² The LCPFCT package solves the generalized continuity equations and has been developed by the US Naval Research Laboratory. Note that the use of the LCPFCT algorithm is advantageous to resolve steep gradients.

The full update to the next time step (\mathbf{W}^{N+1}) requires $\mathbf{W}^{N+1/2}$ and $\mathbf{u}_e^{N+1/2}$, which are not known yet. Because of this, the field solver has to first advance the generalized vorticity \mathbf{W}^N by half a time step (with the input quantities $n_e^{N+1/2}$,

² <http://www.nrl.navy.mil/lcp/LCPFCT>.

$p_e^{N+1/2}$, \mathbf{u}_e^N , \mathbf{W}^N , \mathbf{B}^N) in order to estimate $\mathbf{u}_e^{N+1/2}$, $\mathbf{W}^{N+1/2}$, and $\mathbf{B}^{N+1/2}$, by solving the discretized version of Eq. 9.6:

$$\frac{\partial \mathbf{W}}{\partial t} \Big|_{N \rightarrow N+1/2} = \nabla \times [\mathbf{u}_e^N \times \mathbf{W}^N] - \nabla \times \left(\frac{\nabla p_e^{N+1/2}}{m_e n_e^{N+1/2}} \right). \quad (9.15)$$

A three-dimensional flux-corrected transport algorithm for solving continuity equations (LCPFCT) is used to solve Eq. 9.15 [66].

The solution for $\mathbf{B}^{N+1/2}$ of the discretized version of the elliptic Eq. 9.8 can be obtained by solving

$$\nabla \times \left(\frac{\nabla \times \mathbf{B}^{N+1/2}}{\mu_0 e n_e^{N+1/2}} \right) + \frac{e \mathbf{B}^{N+1/2}}{m_e} = \nabla \times \mathbf{u}_i^{N+1/2} - \mathbf{W}^{N+1/2}. \quad (9.16)$$

The solution of Eq. 9.16 allows to obtain $\mathbf{B}^{N+1/2}$ from $\mathbf{W}^{N+1/2}$. The discretization of this elliptical PDE reveals a system of algebraic equations that can be solved, e.g., by a general purpose scalable linear multigrid solver contained in the multigrid library HYPRE [64, 65].³

Next, the components of $\mathbf{u}_e^{N+1/2}$ can be obtained by solving the discretized Ampère's law, Eq. 9.5.

$$\mathbf{u}_e^{N+1/2} = \mathbf{u}_i^{N+1/2} - \frac{\nabla \times \mathbf{B}^{N+1/2}}{\mu_0 e n_e^{N+1/2}}. \quad (9.17)$$

At this point, all relevant quantities are known at time step $N + 1/2$: $\mathbf{u}_e^{N+1/2}$, $\mathbf{B}^{N+1/2}$, and $\mathbf{W}^{N+1/2}$. These quantities can now be advanced to time step $N + 1$ by solving the following equations:

$$\begin{aligned} \frac{\partial \mathbf{W}}{\partial t} \Big|_{N \rightarrow N+1} &= \nabla \times [\mathbf{u}_e^{N+1/2} \times \mathbf{W}^{N+1/2}] \\ &\quad - \nabla \times \left(\frac{\nabla p_e^{N+1/2}}{m_e n_e^{N+1/2}} \right) \end{aligned} \quad (9.18)$$

$$\nabla \times \left(\frac{\nabla \times \mathbf{B}^{N+1}}{\mu_0 e n_e^{N+1/2}} \right) + \frac{e \mathbf{B}^{N+1}}{m_e} = \nabla \times \mathbf{u}_i^{N+1/2} - \mathbf{W}^{N+1} \quad (9.19)$$

$$\mathbf{u}_e^{N+1} = \mathbf{u}_i^{N+1/2} - \frac{\nabla \times \mathbf{B}^{N+1}}{\mu_0 e n_e^{N+1/2}} \quad (9.20)$$

³ <https://hypre.readthedocs.io/en/latest/ch-intro.html>.

Equation 9.18 can be solved by using already mentioned three-dimensional LCPFCT package developed by the US Naval Research Laboratory and Eq. 9.19 by utilizing the HYPRE library package. This way, one obtains \mathbf{W}^{N+1} , \mathbf{B}^{N+1} , and \mathbf{u}_e^{N+1} . Note that the ion quantities (\mathbf{u}_i and n_i) are used at time steps $(N + 1/2)$, same as for the first half time step. This is justified by the heavy ions whose mass is much larger compared to that of the electrons.

For the calculation of the electric field, Eqs. 9.5, 9.6, and 9.8 are again advanced from time step $N + 1$ to $N + 2$ but using the ion quantities at the time index $N + 1/2$ (i.e., assuming that they are temporally fixed). This reveals \mathbf{B}^{N+2} , \mathbf{W}^{N+2} , and \mathbf{u}_e^{N+2} . Now, the electric field can be calculated at time step $N + 1$ using the discretized version of the generalized Ohm's law, Eq. 9.3:

$$\begin{aligned}\mathbf{E}^{N+1} = & -\frac{m_e}{e} \left[\frac{\mathbf{u}_e^{N+2} - \mathbf{u}_e^N}{2 \Delta t} + (\mathbf{u}_e^{N+1} \cdot \nabla) \mathbf{u}_e^{N+1} \right] \\ & - \mathbf{u}_e^{N+1} \times \mathbf{B}^{N+1} - \frac{\nabla p_e^{N+1}}{en_e^{N+1}}.\end{aligned}\quad (9.21)$$

The value of $p_e^{N+1} = n_e^{N+1} k_B T_e$ is provided by the equation of state, Eq. 9.7, in the same way as before. Note that $p_e^{N+1} = p_e^{N+1/2}$ and $n_e^{N+1} = n_e^{N+1/2}$ are assumed.

Finally, the values of \mathbf{E}^{N+1} and \mathbf{B}^{N+1} are available to move the ions further. This ends a time step and the cycle is repeated.

Note that the model and numerical implementation as presented in Sects. 9.3 and 9.4 do not contain resistive terms. Finite resistivity is usually included in hybrid codes with massless electron fluid to damp whistler waves propagating at the grid scale. This is necessary since the whistler phase velocity for a massless electron fluid increases with the wave number without bounds. As a result, the time steps have to become very small with increasing grid resolution to further satisfy the CFL (Courant) condition and still resolve the whistler wave propagation.

The phase speed of whistlers for an inertial electron fluid, on the other hand, is limited due to the finite electron mass. Therefore, the use of finite resistivity is not necessary for keeping the hybrid codes with massive electrons stable (see Sects. 9.3 and 9.4). It can, however, be implemented in the code to model collisional resistivity in cases of interest. The PIC-typical particle shot noise can be reduced in two different ways. One is the use of higher (at least second) order shape functions for the current density deposition from particles to grid and vice versa to calculate the electromagnetic fields at the ion position. A second possibility in the code is to apply a standard binomial filter on the current deposition, smoothing out the ion current density to be used as a source term in the Maxwell solver.

9.4.4 Code Parallelization and Performance

The CHIEF code is parallelized and optimized for a high efficiency on state-of-the-art supercomputer clusters using a domain decomposition approach implemented based on Message Passing Interface (MPI) commands. The three-dimensional LCPFCT algorithm [66] used to solve Eqs. 9.15 and 9.18 in a massively parallel way is computationally much more efficient than the one-dimensional LCPFCT [67] algorithm used in the nonparallel code (see, e.g., [27]). The scalable linear solvers of the HYPRE package using multigrid methods allow to efficiently solve the elliptic equations 9.16 and 9.19. As a result, it was shown that a massively parallelized CHIEF code performs well for two- as well as for three-dimensional spatial grids. As an example, Fig. 9.1 depicts the strong scaling behavior of the code obtained for quasi-two-dimensional grids on the HPC System COBRA of the Max-Planck-Computing and Data Facility (MPCDF) in Garching (Germany). The COBRA cluster used for the scaling shown in Fig. 9.1 consists of Intel Xeon Gold 6148 (“Skylake” architecture) processors with 40 cores per node, operated at the nominal frequency of 2.4 GHz (Turbo mode disabled). The nodes are connected through a non-blocking 100 Gb/s OmniPath interconnect with a non-blocking fat-tree topology. Further, Intel 19.1.3 C++ and Fortran compilers were used as well as the Intel MPI 2019.7, Intel MKL 2020.4, Blitz++ 1.0.1, and the HYPRE 2.22 libraries. Figure 9.1 shows the run time per time step (average of the first three time

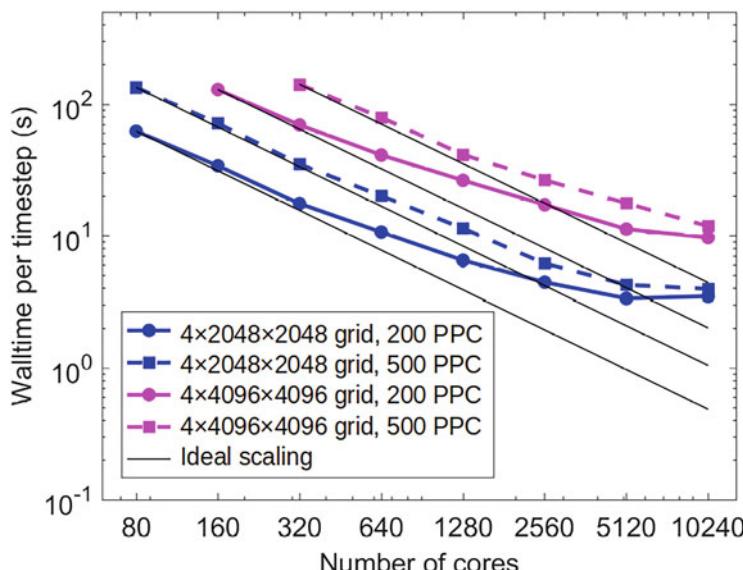


Fig. 9.1 Strong scaling results (run time as a function of compute cores) for the CHIEF code using up to 10,240 cores (256 nodes with 40 compute cores per node). Reprinted from [Jain et al., Physics of Plasmas 29, 053902 (2022)], with the permission of AIP Publishing

steps in four different runs) as a function of the number of computing cores (MPI tasks) for two different grid sizes ($4 \times 2048 \times 2048$, and $4 \times 4096 \times 4096$) up to 10240 cores (256 nodes) with 200 and 500 particles per cell. The black lines indicate ideal strong scaling, taking the computation time on the minimum number of nodes with 768 GB RAM, which could run the calculations as the reference. For the larger grid size ($4 \times 4096 \times 4096$) with 500 particles per cell, the code maintains a good strong scalability, up to roughly 2560 cores. It is able to at least decrease the run time down to ca. 12 s per time step when using 10,240 cores. As expected, deviation from close-to-ideal scaling occurs at smaller core counts for smaller grids or smaller number of particles per cell.

The code CHIEF also performs well on the computer clusters of the Max Planck Institute for Solar System Research in Göttingen (Germany) and the computer clusters of the Technical University Berlin.

9.5 Applications

In the following, we discuss the importance of the consideration of the finite electron mass in hybrid-kinetic codes for four typical space and astrophysical plasma processes: collisionless guide field magnetic reconnection (Sec. 9.5.1), collisionless plasma turbulence (Sec. 9.5.2), collisionless shocks (Sec. 9.5.3), and global simulations of collisionless magnetospheric plasmas (Sec. 9.5.4). In particular, we discuss at some length the importance of the electron inertia for guide field magnetic reconnection by applying the above-described code CHIEF.

9.5.1 Magnetic Reconnection

The magnetic reconnection phenomenon was historically first investigated in the framework of MHD model (see Chap. 1). Later, the consequences of the decoupling of the electron and ion fluids were addressed by Hall-MHD models and simulations (see Chap. 2). The role of kinetic ions was investigated by hybrid-kinetic codes neglecting the electron mass (see Chap. 3). Finally, in the end of the 1990s, the finite electron mass was taken into account in hybrid-kinetic codes to describe magnetic reconnection.

Shay et al. [32], e.g., retained electron inertial terms but only in the equation of the generalized magnetic field and not in that of the generalized electric field [32, 68]. By simulating this way reconnection between two magnetic flux bundles with a zero net-current and a small overlap in order to trigger reconnection, these simulations identified a small region centered around the X-line and extending toward the separatrices of reconnection as a region in which electron inertial effects dominate. These regions with a size of the order of an electron inertial length are much smaller than the Hall region. In these regions, sharp gradients of the electron velocity develop which are susceptible to electron shear flow instabilities.

Kuznetsova et al. [17] retained the convective derivative of the electron fluid velocity and also included the non-gyrotropic electron pressure tensor terms in the

expression of the generalized electric field [17, 34, 35]. Starting with a Harris current sheet equilibrium without an out-of-plane-directed guide magnetic field and adding a small perturbation to trigger reconnection, they found that the electron inertial term is subdued by the pressure tensor term in supporting the steady-state reconnection electric field [17]. Finite electron inertia, however, contributes to the reconnection electric field in a nonsteady fashion for steep gradients at the electron inertial length scale. It was suggested that the electron inertia term may dynamically dominate the influence of the non-gyrotropic pressure tensor terms at electron timescales and for reconnection in finite guide magnetic field.

Almost a decade later, first two-dimensional fully kinetic PIC-code simulations with open boundary conditions showed that reconnection is, indeed, inherently nonstationary due to the development of secondary instabilities of the electron layer [69]. Three dimensional fully kinetic PIC simulations of magnetic reconnection with guide field confirmed the growth of secondary instabilities of the electron current layer [70]. It was shown that the electron inertia is necessary for driving these instabilities [71].

Hybrid-kinetic simulation models with finite electron inertia are able to describe not only the electron-inertia-related instabilities of the thin electron layers as fully kinetic simulations but also their consequences for the larger (ion) scale dynamics of reconnection. The algorithm of the CHIEF code that implements electron inertia without any approximation (see Sec. 9.4) is appropriate for such studies including the description of the electromagnetic fluctuations and turbulence in reconnection regions.

9.5.1.1 Electromagnetic Fluctuations in Reconnection Regions

In-situ space observations of magnetic reconnection have disclosed that electromagnetic and electrostatic fluctuations are generated in and around reconnecting current sheets (see, e.g., [72–74]). Laboratory experiments [23, 75–79] revealed electromagnetic fluctuations in the lower-hybrid frequency range [23, 73, 75] as well as in the ion acoustic [77], Trivelpiece-Gould [76] and Langmuir wave frequency ranges [80]. In the VINETA.II experiments [23], reconnection was driven in a linear laboratory device (cf. the cartoon in Fig. 9.2) by applying an external time varying magnetic field to a plasma immersed in a uniform and constant guide magnetic field B_g in the z (out-of-plane) direction and a figure-eight X-point field $\mathbf{B}_\perp(t)$ in the perpendicular x - y plane. In the resulting reconnecting electron current sheet with a half thickness $\sim 5 d_e \approx \rho_i/2 \approx 25 \rho_e$, electromagnetic fluctuations in the lower-hybrid frequency range were generated. At these scales, ions behave as particles while electrons can be treated as a fluid whose inertia is important. A hybrid-kinetic plasma model with kinetic ions and an inertial electron fluid like CHIEF is, therefore, suitable to describe the fluctuations near the lower-hybrid frequency as observed in space as well as in VINETA.II.

9.5.1.2 EMHD Simulation of Guide Field Magnetic Reconnection with Finite Electron Mass but with Immobile Ions

Before presenting hybrid-PIC CHIEF simulation results of guide field magnetic reconnection, let us demonstrate the role of the electron inertia by first using the framework of an electron-magnetohydrodynamic (EMHD) model. EMHD models are the simplest hybrid models since they allow to study the effects of the electron inertia while neglecting the influence of ions, i.e., considering the ions to be immobile. EMHD simulations of guide field magnetic reconnection were carried out by Jain et al. in 2017 [81] for a configuration similar to that of the VINETA.II experiment [23]. Several consequences of the finite electron mass for guide field reconnection could be revealed already in the framework of the restricted EMHD description: the formation of a thin electron-scale current sheet, development of electron inertial instabilities, and the generation of electromagnetic fluctuations.

The EMHD simulations were initialized with a uniform, motionless plasma embedded in an externally imposed magnetic field $\mathbf{B}^{ext} = \mathbf{B}_{\perp}^{ext} + B_g \hat{z}$ with a guide field $B_g = 15$ mT. The perpendicular magnetic field is created by two infinitely long wires separated by a distance $2d = 30$ cm and carrying a current $I_0 = 2$ kA along the negative z -direction (see Fig. 9.2). To match the experiment geometry, the wire positions are rotated by $\alpha = 30^\circ$ with respect to the y axis. In the experiments, the plasma current is extracted from a localized electron source (the plasma gun) with a radius $r_{gun} = 6$ mm by the large-scale electric field E_z , which in the simulations is approximated by $\mathbf{E}_z^{ext} = -E_{z0} \exp\left(-\frac{x^2+y^2}{2r_{gun}^2}\right) \hat{z}$, with $E_{z0} = 10$ V/m. The simulation box size, 10 cm \times 10 cm \times 10 cm, corresponds to the experimental measurement area. Note that the parallel wires are outside the simulation domain (see Fig. 9.2). The grid resolution in each direction is 1.35 mm. Simulations are run for 3.6 μ s in steps of 0.2 ns. The collisional resistivity is taken to be zero. The simulation results are presented in normalized variables: magnetic fields are normalized by the edge magnetic field produced by the experimental current sheet ($B_{edge} = 0.8$ mT), time by the inverse electron cyclotron frequency in this magnetic field ($\omega_{ce}^{-1} = (eB_{edge}/m_e)^{-1} = 7.2$ ns), distances by the electron inertial length ($d_e = 2.7$ mm), and the current density by $n_0 e v_{Ae} = B_{edge}/\mu_0 d_e = 235.8$ kA/m².

The three-dimensional EMHD simulations have shown that the evolution of the system remains two-dimensional ($\partial/\partial z \approx 0$) during the simulation (electron-) time even for very large simulation boxes in the z -direction. Figures 9.3 and 9.4 show the magnetic fields and currents in an x-y plane. In a small disc around the X-point where the in-plane (perpendicular to the guide magnetic field) magnetic field is vanishingly small, electrons are accelerated by the E_z^{ext} field along the z -direction. Figure 9.5a shows that the current density j_z at the X-point grows until $\omega_{ce} t \approx 100$ and then saturates around $j_z = 0.5 n_0 e v_{Ae} \approx 118$ kA/m², which is of the same order of magnitude as the peak experimental current density (40 kA/m²). The in-plane drift velocity of electrons, primarily given by $\mathbf{v}_{e\perp} = \mathbf{E}_{\perp} \times \hat{z}/B_z$, modifies the initially ($\omega_{ce} t < 10$) disc-shaped cross-section of j_z , respectively stretching and

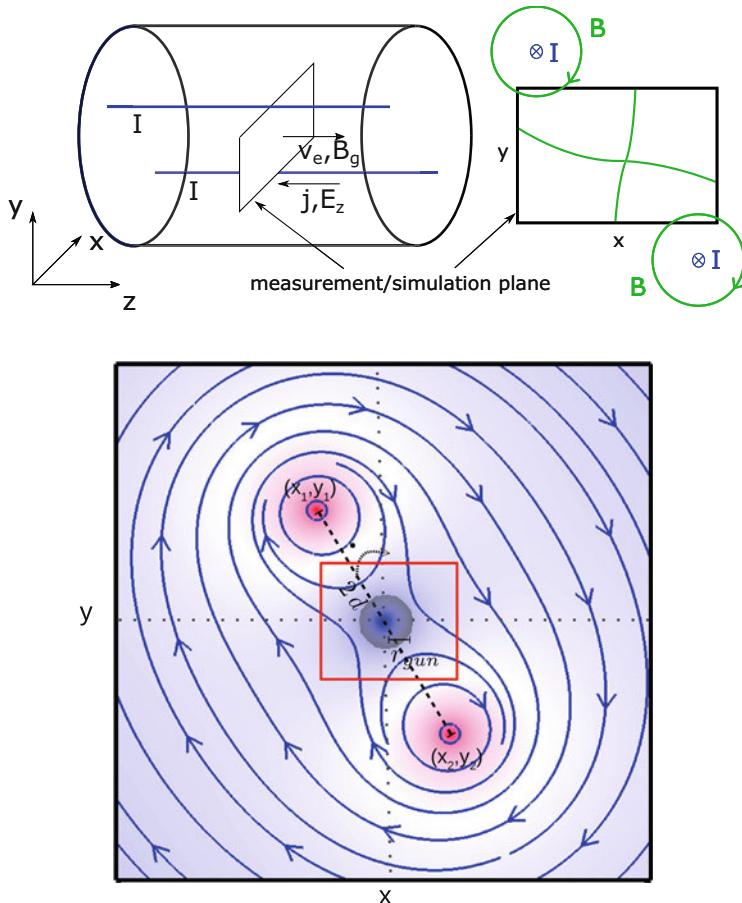


Fig. 9.2 Top: experiment overview and measurement plane. Bottom: simulation plane—color-coded magnitude (red: high and blue: low) and field lines (blue) of the perpendicular external magnetic field \mathbf{B}_{\perp}^{ext} produced by two long wires at the labeled positions (x_1, y_1) and (x_2, y_2) carrying current in the negative z -direction. The shaded area at the center of the simulation box (red square) represents the extent of the external electric field E_z^{ext} (Reproduced from [82] with the permission of AIP Publishing)

pinching the current channel along the directions parallel and perpendicular to the line connecting the conductors. Simultaneously, j_z extends toward the separatrices. The resulting structure of the out-of-plane current sheet at $\omega_{cet} t = 100$ is shown in Fig. 9.3a.

Similar to the experiments, the in-plane electron velocity in the simulations develops a vortical structure. The vortical structure, however, does not align with the current sheet, unlike the experiments. The single vortex formed by $\omega_{cet} t = 60$ (Fig. 9.4a) breaks into two vortices at $\omega_{cet} t = 100$ (Figs. 9.4b and 9.3b) due to the growth of an electron shear flow instability. Linear instability analysis revealed

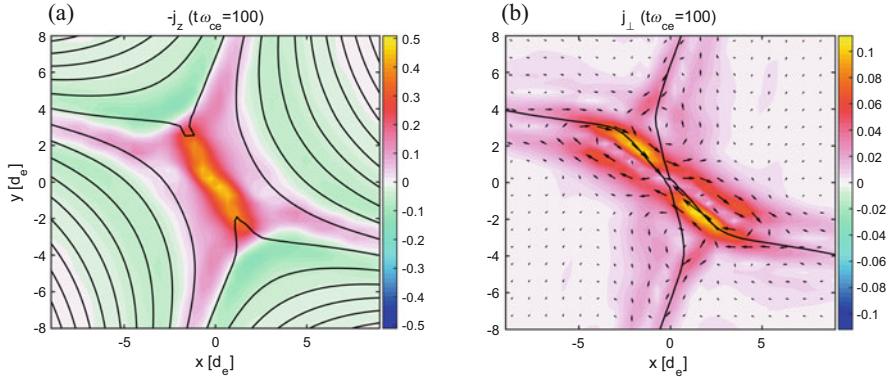


Fig. 9.3 Color-coded axial (a) and in-plane (b) current density with field lines (black) of the total perpendicular magnetic field $\mathbf{B}_{\perp} = \mathbf{B}_{\perp}^p + \mathbf{B}_{\perp}^{ext}$, where \mathbf{B}^p is the magnetic field caused by the plasma currents. Arrows in (b) point in the direction of the in-plane current density (Reproduced from [82] with the permission of AIP Publishing)

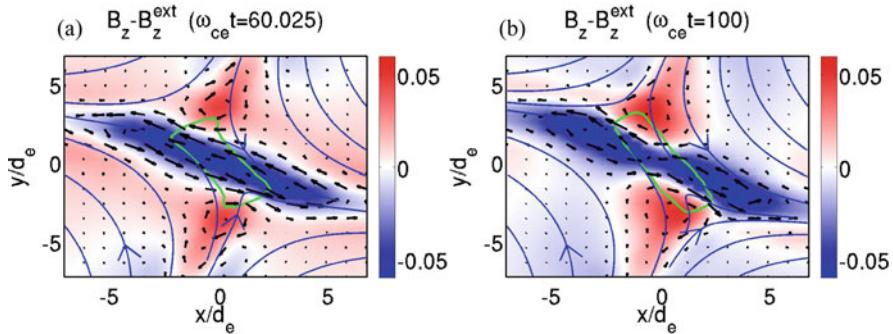


Fig. 9.4 Out-of-plane component of the plasma magnetic field, $\hat{z} \cdot (\mathbf{B} - \mathbf{B}^{ext})$ (color), at two times with projection of magnetic field lines (blue) and electron flow vectors (arrow) in an x-y plane. The green curve is the contour of the axial current density at $j_z = 0.42 j_z^{max}$ (Reproduced from [82] with the permission of AIP Publishing)

that the dominant instability grows on the shear of the in-plane electron flow and requires a finite electron inertia [81]. In the nonlinear state ($\omega_{ce} t > 100$), the instability generates fluctuations in the perpendicular (to the guide field) magnetic field components. The onset of these fluctuations coincides with the saturation of the growth of the axial current density, suggesting that the fluctuations in our simulations with no collisional resistivity can provide some kind of anomalous dissipation (see Fig. 9.5a). EMHD simulations of the same setup without electron inertial terms show that the axial current density at the X-point continues to grow without saturating. Furthermore, the shear flow structures of the out-of-plane and in-plane flows, as shown in Figs. 9.3 and 9.4, did not develop. This suggests that electron inertia plays an important role not only in the development of the instability

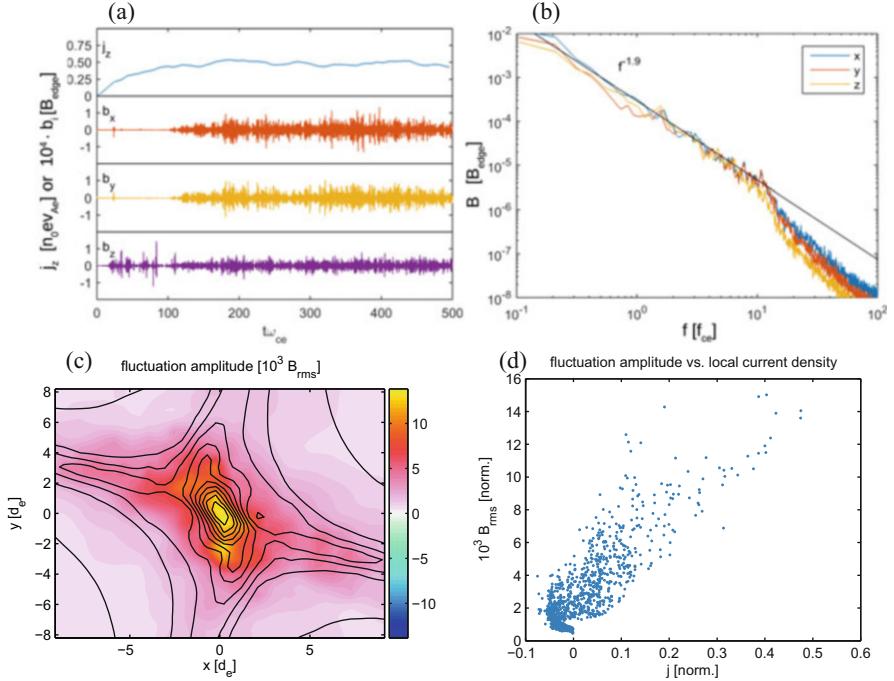


Fig. 9.5 (a) Time evolution of the current density j_z (top panel) and fluctuations in magnetic field components, b_i , ($i = x, y, z$, bottom three panels), obtained by high-pass filtering ($f_{-3\text{dB}} = f_{ce}$) the plasma magnetic field near the X-line. (b) Amplitude spectra of the x , y , and z components of the plasma magnetic field $\mathbf{B}^P = \mathbf{B} - \mathbf{B}^{ext}$ at the X-point over two decades of frequencies ($f/f_{ce} = 0.1 - 10$ or $f/f_{ce,g} \approx 0.005 - 0.5$). (c) Root-mean-square (RMS) fluctuation amplitude (color) and contours of axial current density. (d) Scatter plot of RMS fluctuations amplitudes with the axial current density over space. The Fourier (in b) and RMS (in c and d) amplitudes were obtained from the magnetic field data in the time interval $200 < \omega_{ce} t < 500$. (Reproduced from [82] with the permission of AIP Publishing)

and fluctuations but also in the formation of the shear flow structure on which the instability grows. Note that the electron inertia allows stable simulations even in the absence of resistivity.

The amplitude spectra of the plasma magnetic field fluctuation obtained near the reconnection X-line, Fig. 9.5b, show a clear power law with a spectral index of $\alpha \approx 1.9$ for all the components across two decades of frequencies ($f/f_{ce} = 0.1 - 10$ or $f/f_{ce,g} \approx 0.005 - 0.5$, where $f_{ce} = 2\pi/\omega_{ce}$ and $f_{ce,g} = 2\pi/\omega_{ce,g}$ are the electron cyclotron frequencies in the edge and guide magnetic fields, respectively). The experimental spectra show a broadband power law behavior over a wider frequency range, extending below the lower-hybrid frequency $f_{LH,g}$ in the guide magnetic field. The part of the spectrum at or below $f_{LH,g}$, however, cannot be studied in EMHD simulations since the ion motion is not included in the model. The spectral index $\alpha \approx 1.9$ obtained by EMHD simulations in the high-frequency part

($f > f_{LH,g}$) should be compared to the value $\alpha \approx 2.4$ obtained in the experiment. As in the experiment, the root-mean-square (RMS) values of the magnetic field fluctuations calculated for the frequencies below the electron cyclotron frequency correlate well with the local current density and peak at the center of the current sheet (Fig. 9.5c). Figure 9.5d shows a good correlation between the two quantities with a nearly linear relationship, as found in the experiment.

Hence, the EMHD model reveals already some consequences of the finite electron inertia for guide field magnetic reconnection but fails to describe others, which are associated with the coupling of the electron fluid to the ion motion lasting longer than the typical electron timescales. The inclusion of ion effects requires hybrid-kinetic models with finite electron mass, which consider also the mobility of the ions.

9.5.1.3 Hybrid-Kinetic Simulation of Guide Field Reconnection with Finite Electron Mass and Mobile Ions

A hybrid-kinetic code with an inertial electron fluid like CHIEF allows to investigate the influence of the ions in addition to that of the massive electrons. Let us demonstrate the consequences of mobile ions for guide field reconnection in an initial setup similar (but not identical) to that of EMHD simulations in Sect. 9.5.1.2. Different from the EMHD simulation setup, the spatial position of the current-carrying wires generating the perpendicular magnetic field is now located on the y-axis, i.e., rotated by 30° with respect to their positions in the EMHD simulations discussed in Sect. 9.5.1.2. The other simulation parameters are: $r_{\text{gun}} = 24 \text{ mm}$, $k_B T_e = 6 \text{ eV}$, $T_e/T_i = 60$, ion plasma beta $\beta_i = 0.002$ based on the guide magnetic field $B_g = 8.8 \text{ mT}$, and electron-ion collision frequency $\nu/\Omega_{ce} = 0.01$, with Ω_{ce} the electron cyclotron frequency based on the guide magnetic field. The ratio of the guide magnetic field (B_g) to the upstream magnetic field (B_{edge}) is 11. The simulation box is two-dimensional with a physical size $10 \times 10 \text{ cm}$ and a grid resolution $\Delta x/d_e = 0.25$. The number of particles is 512 protons per cell. A realistic ion-to-electron mass ratio is used. The physical electron inertial length corresponds to $d_e = 0.26 \text{ cm}$.

Figure 9.6 shows that the profile of the out-of-plane current density (panel a) exhibits a slight asymmetry along one separatrix arm that can be attributed to the asymmetric ion density due to the presence of the strong guide field. The in-plane current density in Fig. 9.6b displays a strong current flow along the separatrix. Unlike the EMHD simulations, it is aligned with the current sheet. This is in agreement with the VINETA.II experiment. Similar to the experiments, the power spectrum of the in-plane magnetic fluctuations during a given interval in steady-state ($\Delta t \Omega_{ce} = [12100, 17600]$) in Fig. 9.6c now shows a spectral break near the lower-hybrid frequency Ω_{lh} (based on the guide magnetic field). For frequencies higher than Ω_{lh} , the power spectrum steepens indicating a change in the nature of the electromagnetic fluctuations. The power spectrum has a spectral index $\alpha = 5.5$ for frequencies higher than Ω_{lh} and $\alpha = 2.3$ for frequencies lower than Ω_{lh} . Those values have to be converted to amplitude spectral indices, which results in $\alpha = 1.15$ for $\omega < \Omega_{lh}$ and $\alpha = 2.75$ for $\omega > \Omega_{lh}$. Note that the corresponding experimental

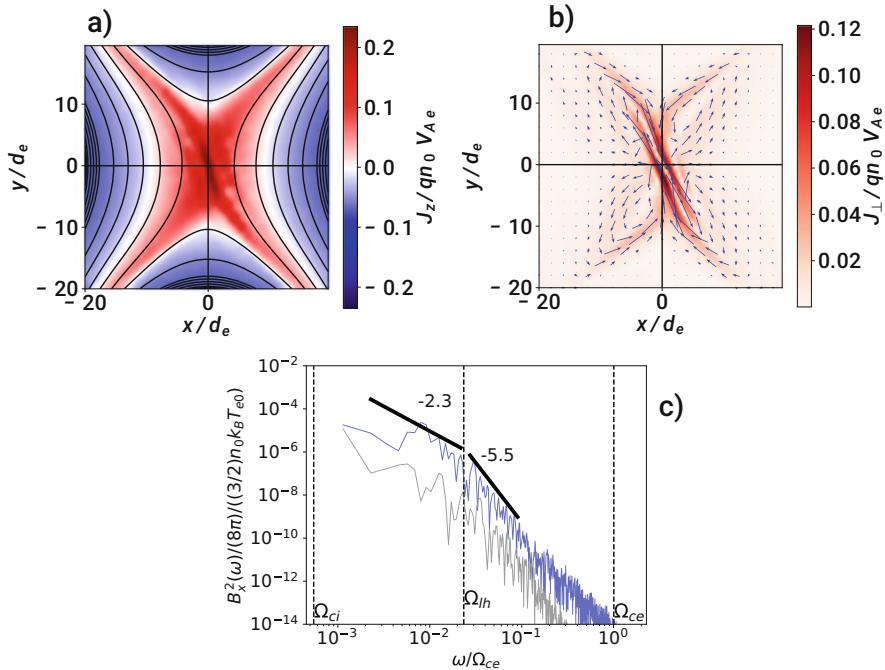


Fig. 9.6 Hybrid-kinetic simulation results for the conditions of the VINETA.II experiment. Contours plots of (a) out-of-plane current density J_z and (b) magnitude of the in-plane current density with the arrows representing the vector flow $\mathbf{J}_\perp = J_x \hat{x} + J_y \hat{y}$. Both quantities were obtained at $t = 12100\Omega_{ce}^{-1}$ (based on the guide magnetic field). The normalization value V_{Ae} is with respect to the upstream magnetic field. (c) Frequency spectrum of in-plane magnetic fluctuations at the X-point (blue lines) and at a point outside of the current sheet (gray line). The oblique straight lines represent an exponential fitting with the indicated spectral slopes below and above the lower-hybrid frequency Ω_{lh}

values are $\alpha = 1.4 - 1.8$ and $\alpha = 2.4$, respectively. Although these values still slightly differ from the experimentally obtained ones, they are closer than those obtained from EMHD simulations due to the consideration of the mobile ions.

9.5.2 Collisionless Plasma Turbulence and Current Sheets

The turbulent transfer of energy from macroscopic to microscopic kinetic scales such as Larmor radii and inertial lengths of plasma particles, where it is finally dissipated into heat, is one of the most viable mechanisms of energy dissipation in collisionless space and astrophysical plasmas. At kinetic scales, current sheets (CSs) with thicknesses ranging from ion to electron scales are observed ubiquitously in space observations of collisionless turbulent plasmas. A number of observational

and simulation studies strongly suggest that the dissipation of energy is localized in and around such kinetic scale CSs.

To understand the dissipative processes in current sheets, two-dimensional hybrid-PIC simulations were performed with massless electrons. These simulations show the formation of ion scale CSs and reveal that they are formed primarily by electron shear flows with an electron bulk flow velocity exceeding by far the ion bulk velocity while density variations are relatively small (<10%) [21].

Current sheets, however, tend to thin down below the ion inertial length. In hybrid-kinetic plasma models neglecting the electron mass, CSs thin down all the way to the numerical grid scale where the thinning is stopped by numerical effects at the grid scale [24]. In real collisionless plasmas, of course, the CS thinning would be stopped by some physical effect at scales below the ion inertial length, like the electron inertial length and/or the electron gyro-radius, a finite-electron-mass effect. Then the CSs may become susceptible to electron shear flow instabilities that can generate the electromagnetic fluctuations observed in space and laboratory experiments (see Sect. 9.5.1.1). The influence of the CS thickness on the turbulence, therefore, has to be studied by taking into account the electron inertia.

Hybrid-kinetic simulations with finite electron mass can be used to study these processes. Let us illustrate this by the results of two-dimensional CHIEF-code simulations initialized with randomly uncorrelated Alfvénic fluctuations,

$$\mathbf{B} = \sum_{k_z=-Nk_z^m/2}^{Nk_z^m/2} \sum_{k_y=-Nk_y^m/2}^{Nk_y^m/2} \frac{B_{rms}\sqrt{2}}{k_{\perp}N} (k_z\hat{\mathbf{y}} - k_y\hat{\mathbf{z}}) \cos(k_z z + k_y y + \phi_b) + B_0\hat{\mathbf{x}} \quad (9.22)$$

$$\mathbf{V} = V_A \sum_{k_z=-Nk_z^m/2}^{Nk_z^m/2} \sum_{k_y=-Nk_y^m/2}^{Nk_y^m/2} \frac{B_{rms}\sqrt{2}}{B_0 k_{\perp} N} (k_z\hat{\mathbf{y}} - k_y\hat{\mathbf{z}}) \cos(k_z z + k_y y + \phi_v), \quad (9.23)$$

where ϕ_b and ϕ_v are independent random phases. For definiteness, an amplitude $B_{rms}/B_0 = 0.24$ and a number of modes $N = 6$ were chosen. The minimum wave numbers allowed in the simulation domain along y- and z-directions are $k_y^m = 2\pi/L_y = 0.061d_i^{-1}$ and $k_z^m = 2\pi/L_z = 0.061d_i^{-1}$, corresponding to the simulation domain length, $L_y = 102.4d_i$ and $L_z = 102.4d_i$, in the two directions, respectively. The maximum perturbed wave numbers along any of the two directions are $0.18d_i^{-1}$. Note that the total magnetic energy of the magnetic perturbation is equal to the bulk flow kinetic energy of the velocity perturbations. A plasma beta based on the out-of-plane background magnetic field of $\beta_i = \beta_e = 0.5$ is chosen and an ion-to-electron mass ratio of $m_i/m_e = 25$. The spatial simulations domain is covered by 1024×1024 grid cells with a grid cell size equal to $\Delta y = \Delta z = 0.5d_e = 0.1d_i$, where d_e (d_i) is the electron (ion) inertial length. Five hundred macro-particles (protons) per cell are launched. The time step is $\Delta t = 0.25\omega_{ce}^{-1} = 0.01\omega_{ci}^{-1}$, where ω_{ce} (ω_{ci}) is

the electron (ion) cyclotron frequency, so that the CFL condition for electron Alfvén wave speed is satisfied to $V_{Ae}\Delta t/\min(\Delta y, \Delta z) = 0.5$. The plasma resistivity is zero.

Inertia-less-electron hybrid-kinetic simulations are also carried out to compare with the results obtained for the inertial electrons with the same parameters except the time step that has to be chosen much smaller (by a factor of five, $\Delta t = 0.002\omega_{ci}^{-1}$) since one now has to stabilize whistler wave instabilities at the grid scale for inertia-less electrons. Boundary conditions are periodic in all directions.

Results are shown in Fig. 9.7. Its two left panels depict iso-contours of the out-of-plane current density obtained for inertia-less (left panel) and for the inertial-electron-fluid (middle panel) simulations. The plots are obtained at $\omega_{ci}B_0t = 120$, the time of maximum turbulence, i.e., when the RMS value of the current density peaks. While the current densities are quite similar in both cases, they are, however, different at the finer scales. This can be seen in their line-outs taken along the line $z/d_i = 25$ (see Fig. 9.7, right panel). While the current densities in both models are more or less the same at larger scales, they differ at shorter scales. The current density for the case of the inertia-less electrons is noisier and spikier than that for obtained with inertial electrons. The magnitude of the current density spikes is smaller if the electron inertia is taken into account. The finite electron inertia interferes with the thinning process of the CSs at scales $k_\perp d_e \sim 1$ before numerical effects (the grid scale) can. This leads to thicker current sheets and smaller current density spikes for inertial electron fluid while the electron bulk flow velocity exceeds the ion bulk velocity in the CSs for both the inertia-less and inertial electron fluid.

Electron inertial effects further contribute to the dissipation of the turbulence via triggering current sheet instabilities. The way the inertial-electron instabilities affect the dissipation in the CSs embedded in collisionless plasma turbulence is, however, still an open question to be investigated.

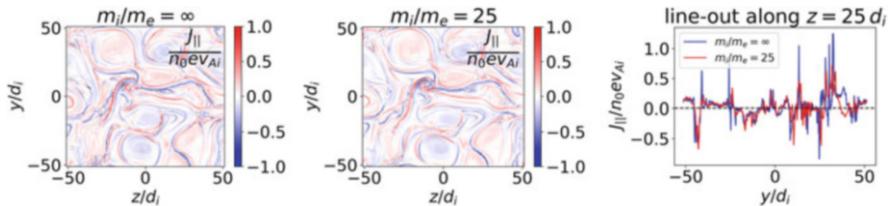


Fig. 9.7 Contours of out-of-plane current density for inertia-less (left panel) and inertial (middle panel) electrons and comparison of their line-outs along $z = 25d_i$ at $t = 120\omega_{ci}^{-1}$, the time of maximum turbulent activity (right panel). Reprinted from [Jain et al., Physics of Plasmas 29, 053902 (2022)], with the permission of AIP Publishing

9.5.3 Collisionless Shocks

So far, collisionless shocks have been simulated mainly by fully kinetic PIC codes or hybrid-kinetic codes with massless electrons. The shock front needs to be sustained by dissipation, which in collisionless plasmas is usually provided by wave-particle interactions due to the plasma instabilities. Hybrid codes with massless electrons do not describe the electron-scale instabilities, which may provide the dissipation in particular shock configurations.

For example, high-Mach number quasi-perpendicular shocks are known to be dominated by whistler waves with wave numbers much larger than the inverse of the ion skin depth and the steep gradient scales. Electron inertial effects were taken into account in hybrid code simulations of shock waves (see, e.g., Chapter 8 of [10]). One of the findings of Ref. [10] for quasi-perpendicular shocks with Mach numbers greater than six is that the periodic reformation of the shock front is accompanied by a standing whistler wave. While the shock front reformation was already described by hybrid codes with massless electrons [83], the formation of standing whistler waves brought in the new physical insights into the breaking of the shock front after its steepening. The whistler waves also cause additional dissipation of the energy of the particles crossing the shock front. Finally, it could be shown this way that the thickness of the shock ramp is comparable with the largest whistler wavelength in the system, which becomes smaller for larger Mach number shocks. Note that hybrid-kinetic shock simulations with electron inertia require, therefore, a relatively fine grid in order to resolve the electron skin depth. Although this makes simulations more expensive, it also allows to smooth out short-wavelength electric fields and stabilize the simulations over longer periods of time compared to standard hybrid code simulations with massless electrons. This effect is even more expressed the larger the mass ratios are.

Another relevant electron inertial effect, also recognized by Ref. [10] (see their Chapter 8), is related to the ion acceleration in low-beta supercritical quasi-perpendicular shocks. In order to reproduce properties of pickup ions as often observed in planetary magnetosphere, an initial ion ring distribution is assumed. Different from the standard diffusive shock acceleration, the shock surfing acceleration is active mainly in quasi-perpendicular shocks. It is due to the formation of a cross-shock potential caused by the different masses of electrons and ions. Ions become trapped between the cross-shock potential and the upstream magnetic field, drifting in this way along the convective electric field becoming energized. This affects mainly the thermal ions, which then become accelerated to the energies that are needed to participate in the standard diffusive shock acceleration. It contributes, therefore, to the solution of the injection problem of the standard diffusive shock acceleration models. A critical ingredient for efficient shock surfing acceleration is a very thin shock front. It was found that for shock surfing acceleration of hydrogen and helium ions, a shock front on the order of the electron skin depth is needed [10]. This finding emphasizes the importance of the electron inertia in shock acceleration processes involving different ion species.

Note that only hybrid codes with electron inertia, or fully kinetic PIC codes which are, however, numerically much more expansive, can reproduce these processes.

9.5.4 Global Magnetospheric Hybrid Code Simulations

After initially global simulations of whole magnetosphere were carried out within the framework of MHD models, the question arose as how to include the physics of magnetopause, magnetotail, reconnection, and shock waves as well as the coupling of magnetosphere and ionosphere. First, hybrid-PIC code simulations were utilized to include ion kinetic effects of reconnection and shock physics but neglecting the electron inertia, for example, by Refs. [8, 84, 85] (see also references therein).

The inclusion of the physics due to the finite electron mass in global simulations is a challenging task that has first been attempted, e.g., by Swift in 1996 [30]. Using curvilinear coordinates, the author considered the effects of the ionospheric plasma by means of a cold-ion-fluid component. His simulations included a term related to the electron inertia into the generalized Ohm's law but neglected it in other places like in the cold-ion-fluid momentum equation. The code completely neglected the contribution of electron pressure. Simulations using this approach revealed that electron inertia has a marginal effect on the code stability compared to the stability provided by the ad hoc assumed resistivity due to the electron-ion friction. It should be noted, however, that in these simulations, the grid cell size always exceeded the electron skin depth.

Later, the same author(s) applied this numerical model to the substorm onset problem, again, taking into account the ionospheric-magnetospheric coupling [31]. They found that the observed wave activity behind the depolarization front could, indeed, be related to the electron inertia, which can cause wave steepening and, finally, contribute to the formation of auroral arcs at electron inertial length scales. The same electron inertia, on the other hand, restricts the maximum accessible current density limiting, this way, the brightness of auroral arcs. It is not clear yet, however, which other effects of the electron inertia might be critical for the global magnetospheric dynamics.

9.6 Future Possible Improvements of Hybrid Code Algorithms

Note that further improvements might make hybrid code algorithms more efficient in the future. The calculation of \mathbf{E}^{N+1} from Eq.(9.21), e.g., requires \mathbf{u}_e^{N+1} and \mathbf{u}_e^{N+2} . The value of \mathbf{u}_e^{N+1} is obtained from Eq.(9.20) using $\mathbf{u}_i^{N+1/2}$ since \mathbf{u}_i^{N+1} is not available, yet, at this point of the calculations. The value of \mathbf{u}_e^{N+2} , on the other hand, is obtained by advancing the equations for \mathbf{W} , \mathbf{B} , and \mathbf{u}_e from time step $N + 1$ to $N + 2$ without updating the ion density and velocity, i.e., by using $\mathbf{u}_i^{N+1/2}$ and n_i^{N+1} . This approach is justified if ions do not move much in a single time step,

which is true for phenomena on space and time scales of the order of or shorter than ion characteristic scales. In fact, correct results for a variety of test problems in this scale range are reproduced by the code CHIEF, which implements the algorithm presented in Sec. 9.4 (see also [27]). If the scales range from electron to ion scales and beyond as it is of interest in certain classes of problems, the method of electric field update needs to be improved. Indeed, the electric field can also be obtained by solving an elliptic PDE, which can be obtained by taking the curl of Faraday's law and using the generalized Ohm's law to substitute for $\partial \mathbf{J} / \partial t$. This equation is,

$$\nabla \times \nabla \times \mathbf{E} + \frac{ne^2}{m_e \epsilon_0 c^2} \left(1 + \frac{m_e}{m_i} \right) \mathbf{E} = -\frac{ne^2}{m_e \epsilon_0 c^2} \left(\mathbf{u}_e + \frac{m_e}{m_i} \mathbf{u}_i \right) \times \mathbf{B} \\ + \frac{e}{m_e \epsilon_0 c^2} \left[m_e \nabla \cdot [n (\mathbf{u}_i \mathbf{u}_i - \mathbf{u}_e \mathbf{u}_e)] + \frac{m_e}{m_i} \nabla \cdot \underline{\mathbf{P}}_i - \nabla p_e \right]. \quad (9.24)$$

Here, $\underline{\mathbf{P}}_i$ is the ion pressure tensor that can be obtained from the second moment of the ion distribution function. The calculation of \mathbf{E}^{N+1} from Eq. 9.24 needs n^{N+1} , \mathbf{u}_e^{N+1} , \mathbf{u}_i^{N+1} , \mathbf{B}^{N+1} , and $\underline{\mathbf{P}}_i^{N+1}$ but not \mathbf{u}_e^{N+2} . This avoids the need of advancing the equations for \mathbf{W} , \mathbf{B} , and \mathbf{u}_e by an additional time step without updating ion quantities. The calculation of \mathbf{E}^{N+1} , however, stays still approximate as \mathbf{u}_i^{N+1} and $\underline{\mathbf{P}}_i^{N+1}$ are not known yet. But now the problem is reduced to the standard issue of hybrid codes, i.e., to the refinement of \mathbf{E}^{N+1} given $\mathbf{u}_i^{N+1/2}$ for which various methods discussed in Chapter 3 of this textbook can be applied, taking into account the additional deposition of the ion pressure tensor onto the grid.

Acknowledgments The work of N.J. was financially supported by the German Science Foundation DFG, projects JA 2680-2-1 and BU 777-16-1, and the work of P.M. was supported by the DFG projects MU-4255/1-1 and BU 777-17-1. We further acknowledge the use of the computing resources of the Max Planck Computing and Data Facility (MPCDF) at Garching, Germany, and the valuable cooperation with Drs. Markus Rampp and Meisam Tabriz supporting the efficient parallelization of the CHIEF code. We also thankfully acknowledge the possibility of using the computing resources of the Max Planck Institute for Solar System Research at Göttingen, Germany, and at the Technical University Berlin.

References

1. R. Treumann, W. Baumjohann, Front. Phys. **1**, 31 (2013). <https://doi.org/10.3389/fphy.2013.00031> <https://www.frontiersin.org/article/10.3389/fphy.2013.00031>
2. R. Bruno, V. Carbone, Living Rev. Sol. Phys. **10**, 2 (2013)
3. R.Z. Sagdeev, Rev. Plasma Phys. **4**, 23 (1966)
4. A. Bret, M.E. Dieckmann, Phys. Plasmas **17**(3), 032109 (2010). <https://doi.org/10.1063/1.3357336>
5. A.B. Langdon, C.K. Birdsall, Phys. Fluids **13**(8), 2115 (1970). <https://doi.org/10.1063/1.1693209> <https://aip.scitation.org/doi/abs/10.1063/1.1693209>
6. A.B. Langdon, Phys. Fluids **22**(1), 163 (1979). <https://doi.org/10.1063/1.862452> <https://aip.scitation.org/doi/abs/10.1063/1.862452>

7. J. Büchner, in *Proceedings of the International School on Space Plasma Simulation*, Kyoto, Feb. 2005, ed. by H. Matsumoto, Y. Omura, H. Usui (Kyoto University Press, Kyoto, 2005), pp. 25–30
8. M. Palmroth, U. Ganse, Y. Pfau-Kempf, M. Battarbee, L. Turc, T. Brito, M. Grandin, S. Hoilijoki, A. Sandroos, S. von Alfthan, *Living Rev. Comput. Astrophys.* **4**, 1 (2018). <https://doi.org/10.1007/s41115-018-0003-2>
9. D. Told, J. Cookmeyer, F. Muller, P. Astfalk, F. Jenko, *New J. Phys.* **18**(6), 065011 (2016). <https://doi.org/10.1088/1367-2630/18/6/065011>
10. A.S. Lipatov, *The Hybrid Multiscale Simulation Technology* (Springer-Verlag, Berlin, 2002). <https://doi.org/10.1007/978-3-662-05012-5>
11. X. Blanco-Cano, N. Omidi, C.T. Russell, *J. Geophys. Res. Space Phys.* **114**(A1) (2009). <https://doi.org/10.1029/2008JA013406>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JA013406>
12. N. Omidi, D. Sibeck, X. Blanco-Cano, D. Rojas-Castillo, D. Turner, H. Zhang, P. Kajdic, *J. Geophys. Res. Space Phys.* **118**(2), 823 (2013). <https://doi.org/10.1002/jgra.50146>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgra.50146>
13. S. Hoilijoki, U. Ganse, D.G. Sibeck, P.A. Cassak, L. Turc, M. Battarbee, R.C. Fear, X. Blanco-Cano, A.P. Dimmock, E.K.J. Kilpua, R. Jarvinen, L. Juusola, Y. Pfau-Kempf, M. Palmroth, *J. Geophys. Res. Space Phys.* **124**(6), 4037 (2019). <https://doi.org/10.1029/2019JA026821>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JA026821>
14. Y. Pfau-Kempf, M. Palmroth, A. Johlander, L. Turc, M. Alho, M. Battarbee, M. Dubart, M. Grandin, U. Ganse, *Phys. Plasmas* **27**(9), 092903 (2020). <https://doi.org/10.1063/5.0020685>
15. N. Omidi, D. Sibeck, O. Gutynska, K.J. Trattner, *J. Geophys. Res. Space Phys.* **119**(4), 2593 (2014). <https://doi.org/10.1002/2013JA019587>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013JA019587>
16. D. Winske, S.P. Gary, *J. Geophys. Res. Space Phys.* **112**(A10) (2007). <https://doi.org/10.1029/2007JA012276>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JA012276>
17. M.M. Kuznetsova, M. Hesse, D. Winske, *J. Geophys. Res.* **103**(A1), 199 (1998). <https://doi.org/10.1029/97JA02699>
18. M. Hesse, D. Winske, *J. Geophys. Res. Space Phys.* **103**(A11), 26479 (1998). <https://doi.org/10.1029/98JA01570>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/98JA01570>
19. A. Le, W. Daughton, H. Karimabadi, J. Egedal, *Phys. Plasmas* **23**(3), 032114 (2016). <https://doi.org/10.1063/1.4943893>
20. L. Franci, S. Landi, L. Matteini, A. Verdini, P. Hellinger, *Astrophys. J.* **812**(1), 21 (2015). <https://doi.org/10.1088/0004-637x/812/1/21>
21. N. Jain, J. Büchner, H. Comișel, U. Motschmann, *Astrophys. J.* **919**(2), 103 (2021). <https://doi.org/10.3847/1538-4357/ac106c>
22. M.S. Weidl, D. Winske, F. Jenko, C. Niemann, *Phys. Plasmas* **23**(12), 122102 (2016). <https://doi.org/10.1063/1.4971231>
23. A. von Stechow, O. Grulke, T. Klinger, *Plasma Phys. Controlled Fusion* **58**(1), 014016 (2016). <https://iopscience.iop.org/article/10.1088/0741-3335/58/1/014016>
24. A.C. Azizabadi, N. Jain, J. Büchner, *Phys. Plasmas* **28**(5), 052904 (2021). <https://doi.org/10.1063/5.0040692>
25. P. Hellinger, P. Travnicek, H. Matsumoto, *Geophys. Res. Lett.* **29**(24), 87 (2002). <https://doi.org/10.1029/2002GL015915>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002GL015915>
26. X. Yuan, I.H. Cairns, P.A. Robinson, *Geophys. Res. Lett.* **34**(2) (2007). <https://doi.org/10.1029/2006GL028447>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL028447>
27. P.A. Muñoz, N. Jain, P. Kilian, J. Büchner, *Comput. Phys. Commun.* **224**, 245 (2018). <https://doi.org/10.1016/j.cpc.2017.10.012>
28. D.W. Forslund, J.P. Freidberg, *Phys. Rev. Lett.* **27**(18), 1189 (1971). <https://doi.org/10.1103/PhysRevLett.27.1189>. <http://link.aps.org/doi/10.1103/PhysRevLett.27.1189>
29. D. Hewett, C. Nielson, *J. Comput. Phys.* **29**(2), 219 (1978). [https://doi.org/10.1016/0021-9991\(78\)90153-5](https://doi.org/10.1016/0021-9991(78)90153-5). <http://linkinghub.elsevier.com/retrieve/pii/0021999178901535>

30. D.W. Swift, J. Comput. Phys. **126**(1), 109 (1996). <https://doi.org/10.1006/jcph.1996.0124>. <http://linkinghub.elsevier.com/retrieve/pii/S0021999196901242>
31. D.W. Swift, Y. Lin, J. Atmos. Sol. Terr. Phys. **63**(7), 683 (2001). [https://doi.org/10.1016/S1364-6826\(00\)00188-7](https://doi.org/10.1016/S1364-6826(00)00188-7). <http://linkinghub.elsevier.com/retrieve/pii/S1364682600001887>
32. M.A. Shay, J.F. Drake, R.E. Denton, D. Biskamp, J. Geophys. Res. **103**(A5), 9165 (1998). <https://doi.org/10.1029/97JA03528>. <http://doi.wiley.com/10.1029/97JA03528>
33. M.A. Shay, J.F. Drake, B.N. Rogers, R.E. Denton, Geophys. Res. Lett. **26**(14), 2163 (1999). <https://doi.org/10.1029/1999GL900481>. <http://doi.wiley.com/10.1029/1999GL900481>
34. M.M. Kuznetsova, M. Hesse, D. Winske, J. Geophys. Res. **105**(A4), 7601 (2000). <https://doi.org/10.1029/1999JA900396>. <http://doi.wiley.com/10.1029/1999JA900396>
35. M.M. Kuznetsova, M. Hesse, D. Winske, J. Geophys. Res. **106**(A3), 3799 (2001). <https://doi.org/10.1029/1999JA001003>. <http://doi.wiley.com/10.1029/1999JA001003>
36. T. Amano, K. Higashimori, K. Shirakawa, J. Comput. Phys. **275**, 197 (2014). <https://doi.org/10.1016/j.jcp.2014.06.048>. <http://linkinghub.elsevier.com/retrieve/pii/S0021999114004550>
37. F. Valentini, P. Trávníček, F. Califano, P. Hellinger, A. Mangeney, J. Comput. Phys. **225**(1), 753 (2007). <https://doi.org/10.1016/j.jcp.2007.01.001>. <http://linkinghub.elsevier.com/retrieve/pii/S0021999107000022>
38. F. Valentini, P. Veltri, F. Califano, A. Mangeney, Phys. Rev. Lett. **101**(2), 025006 (2008). <https://doi.org/10.1103/PhysRevLett.101.025006>. <http://link.aps.org/doi/10.1103/PhysRevLett.101.025006>
39. F. Valentini, F. Califano, P. Veltri, Phys. Rev. Lett. **104**(20), 205002 (2010). <https://doi.org/10.1103/PhysRevLett.104.205002>. <http://link.aps.org/doi/10.1103/PhysRevLett.104.205002>
40. F. Valentini, D. Perrone, P. Veltri, Astrophys. J. **739**(1), 54 (2011). <https://doi.org/10.1088/0004-637X/739/1/54>. <https://iopscience.iop.org/article/10.1088/0004-637X/739/1/54>
41. D. Grošelj, S.S. Cerri, A.B. Navarro, C. Willmott, D. Told, N.F. Loureiro, F. Califano, F. Jenko, Astrophys. J. **847**(1), 28 (2017). <https://doi.org/10.3847/1538-4357/aa894d>
42. F. Califano, S.S. Cerri, M. Faganello, D. Laveder, M. Sisti, M.W. Kunz, Front. Phys. **8**, 317 (2020). <https://doi.org/10.3389/fphy.2020.00317>. <https://www.frontiersin.org/article/10.3389/fphy.2020.00317>
43. S.S. Cerri, S. Servidio, F. Califano, Astrophys. J. **846**(2), L18 (2017). <https://doi.org/10.3847/2041-8213/aa87b0>
44. J. Cheng, S.E. Parker, Y. Chen, D.A. Uzdensky, J. Comput. Phys. **245**, 364 (2013). <https://doi.org/10.1016/j.jcp.2013.03.017>
45. S.T. Jones, S.E. Parker, J. Comput. Phys. **191**(1), 322 (2003). [https://doi.org/10.1016/S0021-9991\(03\)00305-X](https://doi.org/10.1016/S0021-9991(03)00305-X). <http://linkinghub.elsevier.com/retrieve/pii/S002199910300305X>
46. Y.J. Su, S.T. Jones, R.E. Ergun, S.E. Parker, J. Geophys. Res. **109**(A11), A11201 (2004). <https://doi.org/10.1029/2003JA010344>. <http://doi.wiley.com/10.1029/2003JA010344>
47. Y.J. Su, R.E. Ergun, S.T. Jones, R.J. Strangeway, C.C. Chaston, S.E. Parker, J.L. Horwitz, J. Geophys. Res. Space Phys. **112**(A6), A06209 (2007). <https://doi.org/10.1029/2006JA012131>. <http://doi.wiley.com/10.1029/2006JA012131>
48. Y.J. Su, S.T. Jones, R.E. Ergun, F. Bagenal, S.E. Parker, P.A. Delamere, R.L. Lysak, J. Geophys. Res. **111**(A6), A06211 (2006). <https://doi.org/10.1029/2005JA011252>. <http://doi.wiley.com/10.1029/2005JA011252>
49. P. Kilian, T. Burkart, F. Spanier, in *High Performance Computing in Science and Engineering 'II*, ed. by W.E. Nagel, D.B. Kröner, M.M. Resch (Springer, Berlin, 2012), pp. 5–13. <https://doi.org/10.1007/978-3-642-23869-7>
50. N. Jain, Study of Electron-magnetohydrodynamic (EMHD) Phenomena in Plasmas. Ph.D. Thesis, Gujarat University (2006)
51. N. Jain, J. Büchner, Phys. Plasmas **21**, 062116 (2014). <https://doi.org/10.1063/1.4885636>
52. P.A. Muñoz, J. Büchner, Phys. Plasmas **23**(10), 102103 (2016). <https://doi.org/10.1063/1.4963773>
53. N. Jain, J. Büchner, J. Plasma Phys. **81**(06), 905810606 (2015). <https://doi.org/10.1017/S002377815001257>

54. P.A. Muñoz, P. Kilian, J. Büchner, Phys. Plasmas **21**(11), 112106 (2014). <https://doi.org/10.1063/1.4901033>. <http://scitation.aip.org/content/aip/journal/pop/21/11/10.1063/1.4901033>
55. T. Burkart, O. Elbracht, U. Ganse, F. Spanier, Astrophys. J. **720**, 1318 (2010). <https://doi.org/10.1088/0004-637X/720/2/1318>
56. P. Kilian, Teilchenbeschleunigung an kollisionsfreien Schockfronten. Ph.D. thesis (in German), University of Würzburg (2015)
57. U. Ganse, P. Kilian, F. Spanier, R. Vainio, Astron. Astrophys. **564**, A15 (2014). <https://doi.org/10.1051/0004-6361/201322834>
58. C. Schreiner, P. 2, F. Spanier, Astrophys. J. **834**(2), 161 (2017). <https://doi.org/10.3847/1538-4357/834/2/161>
59. K.S. Yee, IEEE Trans. Antennas Propag. **14**(3), 302 (1966). <https://doi.org/10.1109/TAP.1966.1138693>
60. C. Birdsall, A.B. Langdon, *Plasma Physics via Computer Simulation* (IOP Publishing, Bristol, 1991)
61. J.P. Boris, in *Proceedings of the Fourth Conference on the Numerical Simulation of Plasmas*, Washington, ed. by J. Boris (Naval Research Laboratory, Washington, 1970), pp. 3–67
62. R.W. Hockney, J. Eastwood, *Computer Simulation Using Particles* (IOP Publishing, Bristol, 1988)
63. J. Brackbill, J. Comput. Phys. **317**(April), 405 (2016). <https://doi.org/10.1016/j.jcp.2016.04.050>
64. R.D. Falgout, U.M. Yang, in *Computational Science — ICCS 2002*, ed. by P.M.A. Sloot, A.G. Hoekstra, C.J.K. Tan, J.J. Dongarra (Springer, Heidelberg, 2002), pp. 632–641
65. R.D. Falgout, A.J.E. Jones, U.M. Yang, Futur. Gener. Comput. Syst. **22**, 239–251 (2006)
66. S.T. Zalesak, J. Comput. Phys. **31**(3), 335 (1979). [https://doi.org/10.1016/0021-9991\(79\)90051-2](https://doi.org/10.1016/0021-9991(79)90051-2). <http://www.sciencedirect.com/science/article/pii/0021999179900512>
67. J. Boris, A. Landsberg, E. Oran, J. Gardner, LCPFCT – A Flux-Corrected Transport Algorithm for Solving Generalized Continuity Equations. Technical report. NRL Memorandum Report 93-7192 (1993). <http://www.nrl.navy.mil/lcp/LCPFCT>
68. M.A. Shay, J.F. Drake, Geophys. Res. Lett. **25**, 3759 (1998)
69. W. Daughton, J. Scudder, H. Karimabadi, Phys. Plasmas **13**, 072101 (2006)
70. W. Daughton, V. Roytershteyn, H. Karimabadi, L. Yin, B.J. Albright, B. Bergen, K.J. Bowers, Nat. Phys. **7**, 539 (2011)
71. H. Che, J.F. Drake, M. Swisdak, Nature **474**, 184 (2011)
72. J.P. Eastwood, T.D. Phan, B.S. D., A. Tjulin, Phys. Rev. Lett. **102**, 035001 (2009)
73. M. Zhou, X.H. Deng, S.Y. Li, Y. Pang, A. Vaivads, H. Reme, E. Lucek, S. Fu, X. Lin, Z.G. Yuan, W.J. F., J. Geophys. Res. **114**, A02216 (2009)
74. A. Retino, D. Sundkvist, A. Vaivads, F. Mozer, M. Andre, C.J. Owen, Nat. Phys. **3**, 235 (2007)
75. H. Ji, S. Terry, M. Yamada, R. Kulsrud, A. Kuritsyn, Y. Ren, Phys. Rev. Lett. **92**, 115001 (2004). <https://doi.org/10.1103/PhysRevLett.92.115001>. <http://link.aps.org/doi/10.1103/PhysRevLett.92.115001>
76. W. Fox, M. Porkolab, J. Egedal, N. Katz, A. Le, Phys. Plasmas **17**(7), 072303 (2010). <https://doi.org/10.1063/1.3435216>. <http://scitation.aip.org/content/aip/journal/pop/17/7/10.1063/1.3435216>
77. M. Inomoto, A. Kuwahata, H. Tanabe, Y. Ono, T.S. Group, Phys. Plasmas **20**(6) (2013). <https://doi.org/10.1063/1.4811469>. <http://scitation.aip.org/content/aip/journal/pop/20/6/10.1063/1.4811469>
78. S. Dorfman, H. Ji, M. Yamada, J. Yoo, E. Lawrence, C. Myers, T.D. Tharp, Phys. Plasmas **21**(1), 012109 (2014). <https://doi.org/10.1063/1.4862039>. <http://scitation.aip.org/content/aip/journal/pop/21/1/10.1063/1.4862039>
79. A. Kuwahata, H. Igami, E. Kawamori, Y. Kogi, M. Inomoto, Y. Ono, Phys. Plasmas **21**(10), 102116 (2014). <https://doi.org/10.1063/1.4900872>. <http://scitation.aip.org/content/aip/journal/pop/21/10/10.1063/1.4900872>
80. W. Gekelman, R.L. Stenzel, J. Geophys. Res. Space Phys. **89**(A5), 2715 (1984). <https://doi.org/10.1029/JA089iA05p02715>

81. N. Jain, J. Büchner, P.A. Muñoz, Phys. Plasmas **24**, 032303 (2017)
82. N. Jain, A. von Stechow, P.A. Muñoz, J. Büchner, O. Grulke, T. Klinger, Phys. Plasmas **24**(9), 092312 (2017). <https://doi.org/10.1063/1.5004564>
83. K.B. Quest, J. Geophys. Res. **91**(A8), 8805 (1986). <https://doi.org/10.1029/JA091iA08p08805>
84. H. Karimabadi, V. Roytershteyn, H.X. Vu, Y.A. Omelchenko, J. Scudder, W. Daughton, A. Dimmock, K. Nykyri, M. Wan, D. Sibeck, M. Tatineni, A. Majumdar, B. Loring, B. Geveci, Phys. Plasmas **21**(6), 062308 (2014). <https://doi.org/10.1063/1.4882875>. <http://scitation.aip.org/content/aip/journal/pop/21/6/10.1063/1.4882875>
85. H. Karimabadi, H.X. Vu, D. Krauss-Varban, Y. Omelchenko, in *ASP Conference Series*, vol. 359 (2006), pp. 1–7. <http://adsabs.harvard.edu/full/2006ASPC..359..257K>



Generalized Quasi-Neutral Hybrid-Kinetic Simulations

10

Takanobu Amano

Abstract

Understanding the complicated nonlinear space and astrophysical plasma phenomena requires multi-scale and multi-physics numerical modeling. Increasing computational power is allowing us to incorporate additional physics into conventional numerical codes for more physically accurate modeling. We describe a model that generalizes the conventional hybrid algorithm in which ions are fully kinetic whereas electrons are a charge-neutralizing background fluid. Under the assumption of quasi-neutrality, we consider a plasma consisting of electron and ion fluids, as well as arbitrary numbers of fully kinetic species. The electromagnetic field evolves according to Maxwell's equations in the low-frequency approximation. It is shown that the model correctly reduces to Magnetohydrodynamics (MHD), Hall-MHD, Electron-MHD, and conventional hybrid, in appropriate limits. The model substantially extends the applicability of conventional numerical codes and may serve a better alternative.

10.1 Introduction

The quasi-neutrality implies that the charge density in a plasma is close to zero if either appropriate spatial or temporal average is taken. It is often a very good assumption in space and astrophysical plasma applications. Under the condition, we can greatly simplify the physical system to reduce the computational cost because charge-density fluctuations associated with Langmuir waves, which normally have the fastest time scale in plasmas, can be effectively eliminated from the system. The

T. Amano (✉)
University of Tokyo, Tokyo, Japan
e-mail: amano@eps.s.u-tokyo.ac.jp

reduced system yet contains a rich variety of physics that dominates the macroscopic plasma dynamics. This chapter describes a simulation model that takes into account a broad range of physics under the assumption of quasi-neutrality.

One of the motivations for employing such a model for numerical modeling is the continuously increasing computational power. Magnetohydrodynamics (MHD) has been the most widely used model for large-scale plasma phenomena. However, increasing computational resources allow more and more physics to be incorporated into numerical modeling. A natural alternative to the classical MHD is the Hall-MHD model that includes the Hall effect associated with the decoupling between ion and electron dynamics. At the even smaller electron inertial scale, nearly immobile ion dynamics may be ignored, and the so-called Electron-MHD model can be applied. As we see in this chapter, these three fluid models can be combined into a single model, which we call the quasi-neutral two-fluid (QNTF) model [1]. In other words, the QNTF model behaves automatically as the ideal MHD, Hall-MHD, or Electron-MHD, depending on the scale length under consideration.

All the abovementioned models completely ignore kinetic effects such as wave-particle interactions or finite Larmor radius effect, although they often play crucial roles in many applications. For instance, magnetic reconnection is essentially an MHD process in a macroscopic sense, but the kinetic effect plays a key role in breaking the magnetic field frozen-in condition at the diffusion region. Similarly, modeling the nonthermal particle acceleration at a collisionless shock requires the self-consistent description of a macroscopic shock structure and a thin (kinetic-scale) transition layer between the upstream and downstream at the same time.

The classical hybrid model, which also assumes the quasi-neutrality, deals with fully kinetic ions, whereas the electrons are assumed to be a massless charge-neutralizing background fluid. We will show that, by further introducing kinetic populations into the QNTF model, it is extended to what we call the generalized quasi-neutral hybrid (GQNH) model [2]. The model includes the standard numerical models constructed under the assumption of quasi-neutrality (MHD, Hall-MHD, Electron-MHD, classical hybrid) as special cases. It is far more general than these conventional models and has a number of advantages. We will show that it may provide a better alternative for space and astrophysical plasma simulations.

Readers who are not necessarily familiar with the standard numerical models commonly used for numerical plasma modeling are advised to refer to other chapters in the book whenever needed. In particular, a comprehensive review of standard hybrid simulations in Chap. 3 will help to understand the discussion in this chapter. Basic understanding of MHD (Chap. 1) and Hall-MHD (Chap. 2) simulations will also be useful. Electron-MHD and hybrid simulation with finite electron inertia effect are also discussed in Chap. 9. In a sense, the GQNH model tries to combine all these different models together.

In Sect. 10.2, we first remind the reader of the condition in which the quasi-neutrality holds. We then introduce the governing equations in Sect. 10.3 and rewrite them into a form more suitable for numerical treatment including the coupling with kinetic species in Sect. 10.4. In Sect. 10.5, we show that the model correctly reduces to the conventional numerical models in appropriate limits. Section 10.6 briefly

describes numerical methods that are implemented in a code to solve the set of basic equations. A few examples of numerical simulations obtained with the code are shown in Sect. 10.7. Finally, summary and outlook in Sect. 10.8 conclude this chapter.

10.2 Validity of Quasi-neutrality Assumption

First of all, let us show that the quasi-neutrality is a fairly good assumption for a wide range of parameters. One may estimate the magnitude of charge density fluctuations by using Gauss's law $\nabla \cdot \mathbf{E} = 4\pi\varrho$ (where ϱ denotes the charge density). Assuming $E \sim VB/c$ and $|\nabla \cdot \mathbf{E}| \sim E/L$, we obtain the normalized charge density fluctuation:

$$\left| \frac{\delta n}{n_0} \right| \sim \left(\frac{V/L}{\omega_{pe}} \right) \left(\frac{\Omega_{ce}}{\omega_{pe}} \right), \quad (10.1)$$

where V and L are typical velocity and length scales of the phenomena under study. As we will see, the highest frequency in the model is the electron cyclotron frequency Ω_{ce} . This places the limit on the timescale: $1/T \sim V/L \lesssim \Omega_{ce}$ in the above estimate. We thus find that $\omega_{pe}/\Omega_{ce} \gg 1$ is the sufficient condition for the quasi-neutrality. Note that the quasi-neutrality is a good assumption even for $\omega_{pe}/\Omega_{ce} \sim 1$ if sufficiently low-frequency phenomena $\omega_{pe}T \gg 1$ are considered.

It is easy to confirm that the longitudinal (or curl-free) part of the displacement current in Maxwell's equations is related to the charge density fluctuations via the charge continuity equation. On the other hand, the transverse (or divergence-free) part of the displacement current describes the electromagnetic waves above the cutoff frequency ($\sim \omega_{pe}$). Therefore, we can safely ignore the displacement current for sufficiently low-frequency phenomena ($\omega_{pe}T \gg 1$), which is indeed physically consistent with the assumption of quasi-neutrality.

Therefore, we conclude that the quasi-neutrality and the low-frequency approximation to Maxwell's equations will hold quite well, as long as the timescale of interest is much longer than the inverse plasma frequency.

10.3 Governing Equations

We may write Maxwell's equations as follows by ignoring the displacement current:

$$\frac{1}{c} \frac{\partial}{\partial t} \mathbf{B} = -\nabla \times \mathbf{E} \quad (10.2)$$

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{J}, \quad (10.3)$$

$$\nabla \cdot \mathbf{B} = 0. \quad (10.4)$$

$$\nabla \cdot \mathbf{E} = 4\pi\varrho. \quad (10.5)$$

Note that the quasi-neutrality implies that $\varrho \sim 0$ is always satisfied.

For the moment, we consider the system consisting of arbitrary numbers of fluid and kinetic species, which are respectively governed by the Euler and the Vlasov equations. The total charge and current densities (ϱ, \mathbf{J}) should be calculated by taking the sum of contributions overall charged-particle species in the system. For convenience, we divide them into contributions from fluid species and kinetic species and formally write

$$\varrho = \varrho_f + \varrho_k \quad (10.6)$$

$$\mathbf{J} = \mathbf{J}_f + \mathbf{J}_k \quad (10.7)$$

where ϱ_f and ϱ_k denote charge densities for fluid and kinetic species. Similarly, \mathbf{J}_f and \mathbf{J}_k are current densities for fluid and kinetic species.

We assume that the force per unit mass acting on a particle (with a charge q , mass m , moving with velocity \mathbf{v}) is given by the sum of Lorentz force and a yet unspecified friction \mathbf{R}

$$\mathbf{F} = \frac{q}{m} \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right) + \mathbf{R}. \quad (10.8)$$

The dynamics of all the particle species are coupled through the Lorentz force (and the frictional force). Other external forces such as gravity can easily be incorporated if necessary.

The Euler equations for a fluid species denoted by s are given by:

$$\frac{\partial}{\partial t} \rho_s + \nabla \cdot (\rho_s \mathbf{u}_s) = 0, \quad (10.9)$$

$$\frac{\partial}{\partial t} \rho_s \mathbf{u}_s + \nabla \cdot (\rho_s \mathbf{u}_s \mathbf{u}_s + p_s \mathbf{I}) = \rho_s \mathbf{F}_s, \quad (10.10)$$

$$\frac{\partial}{\partial t} \varepsilon_s + \nabla \cdot \{(\varepsilon_s + p_s) \mathbf{u}_s\} = \rho_s \mathbf{F}_s \cdot \mathbf{E} \quad (10.11)$$

where $q_s, m_s, \rho_s, \mathbf{u}_s, p_s, \varepsilon_s$ are the charge, mass, mass density, bulk velocity, and (scalar) pressure, and total fluid energy density for the particle species s , respectively. Note that \mathbf{I} denotes the identity tensor. We assume a polytropic equation of state with a specific heat ratio γ , for which we have:

$$\varepsilon_s = \frac{1}{2} \rho_s \mathbf{u}_s^2 + \frac{1}{\gamma - 1} p_s. \quad (10.12)$$

The Vlasov equation for a kinetic species s is given by

$$\frac{\partial}{\partial t} f_s + \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{x}} f_s + \mathbf{F} \cdot \frac{\partial}{\partial \mathbf{v}} f_s = 0, \quad (10.13)$$

where $f_s(\mathbf{x}, \mathbf{v}, t)$ is the phase space density (PSD) multiplied by the mass of species m_s so that the zeroth-order velocity moment gives the mass density rather than number density. Note that the factor m_s is not usually included in the definition of PSD, but this makes notations in the following more consistent.

The charge and current densities for the fluid and kinetic species are given by

$$\varrho_k = \sum_{s=kinetic} \frac{q_s}{m_s} \int f_s(\mathbf{v}) d\mathbf{v} \quad (10.14)$$

$$\mathbf{J}_k = \sum_{s=kinetic} \frac{q_s}{m_s} \int f_s(\mathbf{v}) \mathbf{v} d\mathbf{v}. \quad (10.15)$$

$$\varrho_f = \sum_{s=fluid} \frac{q_s}{m_s} \rho_s \quad (10.16)$$

$$\mathbf{J}_f = \sum_{s=fluid} \frac{q_s}{m_s} \rho_s \mathbf{u}_s. \quad (10.17)$$

We should note that no assumption has been made so far on the number of species, charge-to-mass ratio, mass density, energy density, etc., both for the fluid and kinetic species. Therefore, it offers a quite general set of equations that is correct as long as the assumption of quasi-neutrality holds. Various limiting cases will be discussed later in Sect. 10.5.

10.4 Quasi-Neutral Two Fluids Coupled with Kinetic Populations

When we try to solve the governing equations, it is immediately found that we cannot deal with all the fluid variables (for fluid species) and the PSDs (for kinetic species) as independent variables simultaneously. This is because otherwise, a finite charge density perturbation would arise in general, which violates the quasi-neutrality assumption. This is the fundamental difference from the system described by the full set of Maxwell's equations (i.e., with the displacement current term). In the presence of a finite charge density, the displacement current let the electric field evolve in time, which generates a Langmuir wave. Therefore, all the fluid and kinetic species can evolve according to their own governing equations.

In the absence of the displacement current, Gauss's law and Ampere's law place constraints on the system. The density for one of the species should be determined to satisfy the charge neutrality $\varrho = 0$. Similarly, the total current density given by the

magnetic field $\nabla \times \mathbf{B}$ automatically determines the bulk velocity of the species. In conventional Hall-MHD or hybrid codes, the electrons are assumed to be a massless fluid and its density and bulk velocity are determined by the constraints. Although we also determine the fluid electron density and bulk velocity in the same way, as will be shown below, the electron inertia is not necessarily small.

To proceed further, we restrict ourselves to the system that contains only two fluid species, namely, electrons and ions $s = i, e$. (This restriction is not absolutely necessary but introduced here just to simplify the model; one may easily extend the model to take into account multi-fluid effects, which is important for some applications.) Instead of solving the original Euler equations separately for the two fluids with the source terms on the right-hand side, we rewrite the basic equations for the fluid part into a form similar to the MHD equations. Taking the sum of the two fluid equations, we may obtain the following equations in the conservative form:

$$\frac{\partial}{\partial t} \left[\sum_{s=i,e} \rho_s \right] + \nabla \cdot \left[\sum_{s=i,e} \rho_s \mathbf{u}_s \right] = 0. \quad (10.18)$$

$$\frac{\partial}{\partial t} \left[\sum_{s=i,e} \rho_s \mathbf{u}_s \right] + \nabla \cdot \left[\sum_{s=i,e} (\rho_s \mathbf{u}_s \mathbf{u}_s + p_s \mathbf{I}) + \frac{\mathbf{B}^2}{8\pi} \mathbf{I} - \frac{\mathbf{B}\mathbf{B}}{4\pi} \right] = \mathbf{F}_{\text{ext}}, \quad (10.19)$$

$$\begin{aligned} & \frac{\partial}{\partial t} \left[\sum_{s=i,e} \varepsilon_s + \frac{\mathbf{B}^2}{8\pi} \right] \\ & + \nabla \cdot \left[\sum_{s=i,e} (\varepsilon_s + p_s) \mathbf{u}_s + c \frac{\mathbf{E} \times \mathbf{B}}{4\pi} \right] = Q_{\text{ext}}. \end{aligned} \quad (10.20)$$

The left-hand sides are understood as the conservation laws for the mass, momentum, and energy, respectively. The source terms on the right-hand side are given by

$$\mathbf{F}_{\text{ext}} = - \left(\varrho_{\mathbf{k}} \mathbf{E} + \frac{\mathbf{J}_{\mathbf{k}}}{c} \times \mathbf{B} \right) + \sum_{s=i,e} \rho_s \mathbf{R}_s \quad (10.21)$$

$$Q_{\text{ext}} = - \mathbf{J}_{\mathbf{k}} \cdot \mathbf{E} + \sum_{s=i,e} \rho_s \mathbf{u}_s \cdot \mathbf{R}_s. \quad (10.22)$$

The first terms describe the effect of the Lorentz force on the kinetic species that appear here because of the conservation of total momentum or energy for the entire system. In other words, the momentum and energy gained (lost) by the kinetic species are compensated by the corresponding loss (gain) by the fluid species. The

second terms arise due to the frictional force. Later, we will introduce a frictional force that satisfies the total momentum conservation law so that the frictional momentum change in the fluid species is compensated by the kinetic species. The frictional effect does, however, remain in the energy equation as irreversible heating. We should mention that the conservative form is particularly useful for numerical treatment as, for instance, advanced shock-capturing schemes for the ideal MHD equations can relatively easily be applied.

Notice that the above conservation laws involve both the electron and ion quantities. However, the electron density and bulk velocity may be determined using the constraints imposed by Gauss's law and Ampere's law. Denoting the conservative variables: $D = \rho_i + \rho_e$, $\mathbf{M} = \rho_i \mathbf{u}_i + \rho_e \mathbf{u}_e$, $K = \varepsilon_i + \varepsilon_e + \mathbf{B}^2/8\pi$, we obtain the following relations

$$\rho_e = \frac{\varrho_k + \frac{q_i}{m_i} D - \frac{1}{4\pi} \nabla \cdot \mathbf{E}}{\frac{q_i}{m_i} - \frac{q_e}{m_e}}, \quad (10.23)$$

$$\mathbf{u}_e = \frac{\mathbf{J}_k + \frac{q_i}{m_i} \mathbf{M} - \frac{c}{4\pi} \nabla \times \mathbf{B}}{\rho_e \left(\frac{q_i}{m_i} - \frac{q_e}{m_e} \right)}, \quad (10.24)$$

$$p_i + p_e = (\gamma - 1) \left(K - \frac{1}{2} \rho_i \mathbf{u}_i^2 - \frac{1}{2} \rho_e \mathbf{u}_e^2 - \frac{\mathbf{B}^2}{8\pi} \right). \quad (10.25)$$

Since D and \mathbf{M} are obtained by updating the conservation laws, we see that the density and bulk velocity both for fluid electrons and ions can be calculated separately provided that the electromagnetic field is given. On the other hand, it is only the total fluid pressure ($p_i + p_e$) that can be obtained from the conservation laws. We thus need a recipe to determine the individual pressure contributions that are needed for the generalized Ohm's law as introduced below. The most straightforward way is to obtain the electron pressure by solving equation (10.11) for the electron fluid. Since the electron pressure effect is often unimportant for space plasma applications, simpler methods are often adopted in typical hybrid codes. For instance, use of $p_e \propto \rho_e^\gamma$ is one of the simplest choices. However, this does not take into account irreversible heating (resistive and viscous heating at shocks or sharp discontinuities). Alternatively, one may introduce an ion to electron temperature ratio $\tau = T_i/T_e$ and assume it to be a constant [1]. In any case, one needs to adopt a model for irreversible heating because the energy partition between ions and electrons is, in general, not a trivial problem.

Maxwell's equations in the low-frequency approximation do not include a temporal derivative for the electric field. Therefore, we have to use Ohm's law to determine the electric field from known quantities. In typical hybrid or Hall-MHD codes, an approximated form of the generalized Ohm's law obtained from the equation of motion for the electron fluid is used. However, this approximation is not necessary. It is actually possible to obtain an exact form that is fully consistent with the original governing equations. This may be done by multiplying q_s/m_s to

the equation of motion for the two fluids and taking the sum, yielding (see [1,2] for derivation):

$$\left(\Lambda + c^2 \nabla \times \nabla \times \right) \mathbf{E} = - \frac{\boldsymbol{\Gamma}}{c} \times \mathbf{B} + \nabla \cdot \boldsymbol{\Pi} + \eta \Lambda \mathbf{J}. \quad (10.26)$$

Here, we have introduced a phenomenological resistivity η for the sake of numerical convenience, although, in most cases, we ignore the resistivity $\eta = 0$. The moment quantities $\Lambda = \Lambda_f + \Lambda_k$, $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_f + \boldsymbol{\Gamma}_k$, $\boldsymbol{\Pi} = \boldsymbol{\Pi}_f + \boldsymbol{\Pi}_k$ in the above equation are defined as follows

$$\Lambda_k = \sum_{s=kinetic} \frac{4\pi q_s^2}{m_s^2} \int f_s d\mathbf{v}, \quad (10.27)$$

$$\boldsymbol{\Gamma}_k = \sum_{s=kinetic} \frac{4\pi q_s^2}{m_s^2} \int \mathbf{v} f_s d\mathbf{v}, \quad (10.28)$$

$$\boldsymbol{\Pi}_k = \sum_{s=kinetic} \frac{4\pi q_s}{m_s} \int \mathbf{v} \mathbf{v} f_s d\mathbf{v}, \quad (10.29)$$

$$\Lambda_f = \sum_{s=i,e} \frac{4\pi q_s^2}{m_s^2} \rho_s, \quad (10.30)$$

$$\boldsymbol{\Gamma}_f = \sum_{s=i,e} \frac{4\pi q_s^2}{m_s^2} \rho_s \mathbf{u}_s \quad (10.31)$$

$$\boldsymbol{\Pi}_f = \sum_{s=i,e} \frac{4\pi q_s}{m_s} (\rho_s \mathbf{u}_s \mathbf{u}_s + p_s \mathbf{I}). \quad (10.32)$$

These quantities may be easily calculated from the fluid moments and PSDs. Therefore, the electric field can be determined from the generalized Ohm's law. To obtain the resistive term in a consistent manner, we adopt the frictional force of the form:

$$\mathbf{R}_s = - \frac{1}{4\pi} \eta \Lambda (\mathbf{u}_s - \mathbf{U}), \quad (10.33)$$

where

$$\mathbf{U} = \frac{\sum_s \rho_s \mathbf{u}_s}{\sum_s \rho_s} \quad (10.34)$$

is the center-of-mass velocity. The summation here must be taken over all species including kinetic populations, for which ρ_s , \mathbf{u}_s must be understood as corresponding moments. It is easy to confirm that the frictional force satisfies the conservation law for the total momentum. We should note that the generalized Ohm's law is exact in the sense that it is mathematically equivalent to the original governing equations. We also note that essentially the same Ohm's law (although slightly less generic) has been obtained by Shumaker et al. [3] and Lipatov [4].

The basic equations for the GQNH model may be summarized as follows. The conservation laws (10.18)–(10.20) and the Vlasov equation (10.13) describe the evolution for the fluid and the kinetic parts, respectively. The electron pressure can be determined by solving the energy equation for the electron fluid equation (10.11) or by using a simpler equation of state. The moment quantities are then used to determine the electric field via the generalized Ohm's law (10.26). With the electric field, the magnetic field can be updated using Faraday's law (10.2).

10.5 Limiting Cases

The GQNH model offers a much more general numerical model than the conventional codes. Therefore, it is instructive to discuss the physics included in the model by showing that it naturally reduces to standard plasma simulation models when appropriate limits are taken.

10.5.1 Quasi-Neutral Two-Fluid Model

Let us first consider the QNTF model in which the kinetic populations are absent. It reduces to the classical MHD in the long wavelength limit, whereas the ion and electron inertial effects become progressively important at short wavelengths comparable to the ion and electron inertial lengths, respectively. This may be seen by investigating the linear dispersion relation obtained for a homogeneous plasma (see [1] for detailed derivation):

$$P\tilde{\omega}^6 - Q\tilde{\omega}^4 + R\tilde{\omega}^2 - S = 0 \quad (10.35)$$

with the coefficients defined as follows

$$P = \left(1 + \epsilon\tilde{k}^2\right)^2, \quad (10.36)$$

$$Q = 1 + \beta + (1 + \tilde{k}^2)\cos^2\theta + \epsilon\tilde{k}^2\left(1 - \cos^2\theta + 2\beta\right) + \epsilon^2\tilde{k}^2\left(\cos^2\theta + \beta\tilde{k}^2\right), \quad (10.37)$$

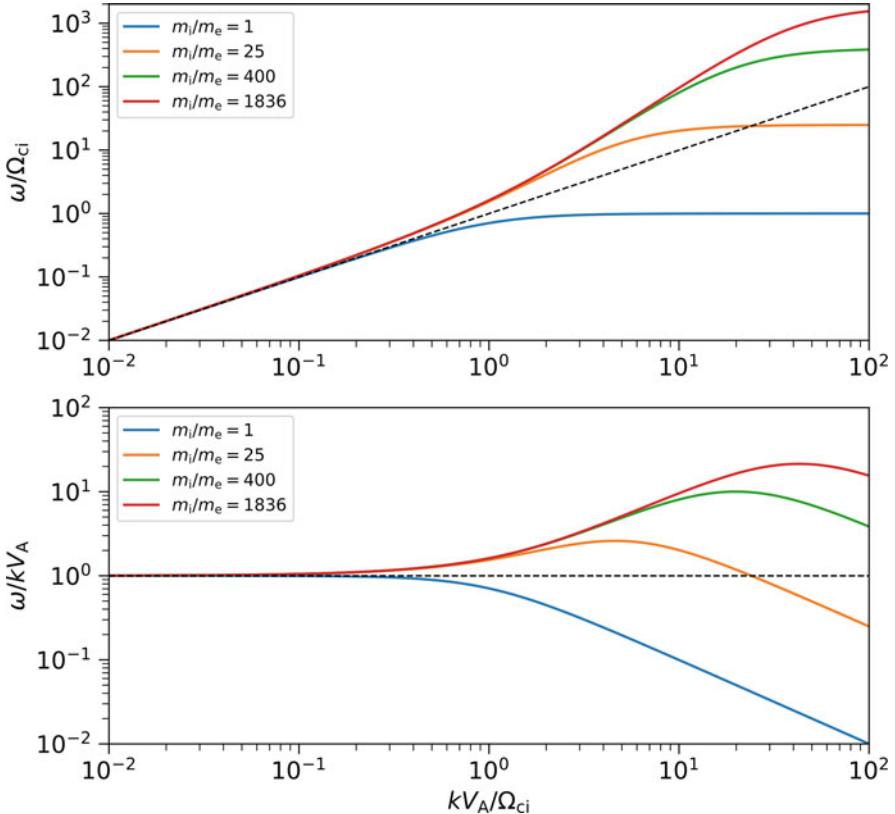


Fig. 10.1 Frequency (top) and phase speed (bottom) of right-hand polarized waves propagating parallel to the ambient magnetic field as functions of wavenumber for four different mass ratios $m_i/m_e = 1, 25, 400, 1836$. The black dashes lined represent the long wavelength (or ideal MHD) limit

$$R = \left(1 + (2 + \tilde{k}^2)\beta + \epsilon^2 \tilde{k}^2 \beta\right) \cos^2 \theta, \quad (10.38)$$

$$S = \beta \cos^4 \theta, \quad (10.39)$$

where $\tilde{\omega} = \omega/kV_A$ and $\tilde{k} = kV_A/\Omega_{ci}$ are the normalized phase speed and wavenumber. Here, we define the Alfvén speed $V_A = B/\sqrt{4\pi(\rho_i + \rho_e)}$, ion cyclotron frequency $\Omega_{ci} = q_i B/m_i c$, the electron-to-ion mass ratio $\epsilon \equiv m_e/m_i$, and the wave propagation angle with respect to the ambient magnetic field θ .

Note that the dispersion relation is a sixth-order polynomial equation, which is the same as the MHD. It is easy to confirm that the dispersion relation becomes identical to the ideal MHD in the limit $\tilde{k} \rightarrow 0$ and $\epsilon \rightarrow 0$. A finite \tilde{k} represents the ion inertial (or Hall) effect. Similarly, a finite ϵ introduces the electron inertial effect. These properties are confirmed in Fig. 10.1, showing the real frequency and phase

velocity of right-hand polarized waves propagating parallel to the ambient magnetic field. Corresponding plots for the left-hand polarized mode are nearly identical to $m_i/m_e = 1$ and are not shown. We see that the deviation from the ideal MHD limit $kV_A/\Omega_{ci} \rightarrow 0$ (shown in black dashed lines) first appears at around $kV_A/\Omega_{ci} \sim 1$ due to the ion inertial effect. The decoupling of ions and electron dynamics increases the phase speed of right-hand polarized waves for $m_i/m_e > 1$, which then peaks at around $kV_A/\Omega_{ci} \approx \sqrt{m_i/m_e}$ due to the electron inertial effect. Note that the reduction of phase speed due to the finite electron inertia effect tends to stabilize the code because otherwise, the phase speed increases in proportion to wavenumber without bounds. This is indeed one of the reasons why it is difficult to use a fine mesh in Hall-MHD and/or hybrid codes.

It is important to note that the model correctly reproduces the dispersion relation even for $m_i/m_e = 1$, i.e., an electron-positron plasma. This is natural as we have not introduced any assumptions on the electron inertia. In this case, the Hall effect disappears due to the symmetry between the positive and negative charges. It is easy to check the dispersion relation for right-hand and left-hand polarized waves are identical. The phase speed at short wavelength decreases from the ideal MHD, potentially making a simulation code numerically more stable.

10.5.2 Effect of Kinetic Populations

Now, we consider the system in which the kinetic ions provide the dominant (positive) charge density and the fluid ion density is negligible. The charge neutrality is of course maintained by the presence of the electron fluid. In other words, this limit essentially corresponds to the standard hybrid model, although approximated forms of Ohm's law are often employed in typical hybrid codes. Use of the exact generalized Ohm's law (10.26) has both numerical and physical advantages.

First, as we have seen in the previous subsection, the finite electron inertial effect is fully taken into account in the model. The method to incorporate this effect into the code differs from the conventional hybrid codes [4] (see also Chap. 9). It is customary to include finite electron inertia correction terms into the electric and magnetic fields (though the correction to the electric field is often ignored). The corrections can be included into an existing hybrid code relatively easily without changing the structure of the code.

On the other hand, the electric field obtained via the generalized Ohm's law (10.26) includes the correction from the first place, which is then used to advance the magnetic field via the induction equation. The advantage of using this method is that it allows the simulation box to contain a pure vacuum region [5]. This can be seen by taking the limit of negligible density, which implies $\Lambda \rightarrow 0$, $\Gamma \rightarrow 0$, $\Pi \rightarrow 0$. The generalized Ohm's law (10.26) then reduces to

$$\nabla^2 \mathbf{E} = 0, \quad (10.40)$$

where $\nabla \cdot \mathbf{E} = 0$ is used. We see that the electric field in a vacuum region is given by the solution of Laplace's equation [5, 6]. This is in clear contrast to conventional hybrid codes, in which the zero-density limit is intrinsically undefined because the electric field is always obtained via division by density. The use of the exact Ohm's law eliminates this operation. Appearance of low-density (near vacuum) regions is one of the reasons for the conventional hybrid codes become numerically unstable. It is a desirable property in constructing a more robust code [5]. We should note, however, that there exist different approaches for handling a vacuum region in hybrid simulations. For instance, the vacuum region may be replaced by a resistive medium [7, 8].

We now consider the case in which both the fluid ions and kinetic populations contribute to the dynamics. One of the motivations to consider such a model is to describe the self-consistent coupling between nonthermal energetic particles (EPs) and the MHD fluid. The EP population usually has a negligible density, but non-negligible or even dominant contribution to the energy density. Therefore, it is crucial to take into account the coupling between them, preferably with the kinetic effect included, even for macroscopic dynamics. We should mention that the scale length in which the kinetic effect becomes important is given by the typical gyroradius, which can be much larger than the ion inertial length and/or the gyroradius of thermal ions. At this length scale, the ideal MHD should be an adequate model if EPs are absent. It is thus reasonable to adopt the fluid approximation for the thermal component, whereas EPs should be a fully kinetic species. Such a model is referred to as an Energetic-Particle-MHD (EP-MHD) hybrid model.

The GQNH model includes the EP-MHD hybrid as a special case. It automatically reduces to the ideal MHD in the long wavelength limit and in the absence of EPs. The kinetic effect associated with a finite density EP population (such as wave-particle interaction and finite Larmor radius effect) is fully taken into account. It is worth noting that similar models have been proposed to study the self-consistent coupling between cosmic rays (CRs) and the MHD fluid for astrophysical applications [9, 10]. The advantage of GQNH over the existing codes is more extended applicability. We will see that the kinetic effect of energetic electrons is correctly included only when the exact form of Ohm's law is used. In addition, a simulation box may contain regions with very different parameter regimes at the same time. For instance, one essentially described by the ideal MHD and another by the standard hybrid may coexist without any artificial boundaries between them. This will be a useful property for numerical modeling of complicated systems, where regions of vastly different parameter regimes coexist in a single simulation box.

10.6 Numerical Methods

As we have seen so far, the GQNH model contains a lot more physics than the conventional hybrid codes. Therefore, a single code that numerically solves the set of basic equations can apply to a wide range of problems. In the following, we present a brief summary of numerical methods that have successfully been implemented in a parallelized three-dimensional (3D) code [2]. The description, however, does not mean to be comprehensive. In particular, since the numerical scheme is designed genuinely for multidimensional problems and requires basic understanding of shock-capturing schemes, we feel it is not necessarily a good idea to present the details. Interested readers may refer [1, 2] for more complete descriptions and discussion. We also note that the method presented here is just an example that works reasonably well. There is certainly room for further improvement, which is left for the readers.

The basic equations are roughly divided into the fluid and kinetic parts. The formal structure of equations for the fluid part is essentially the same as the MHD: the conservation laws and the magnetic induction equation supplemented by Ohm's law that determines the electric field. The effects of kinetic populations are contained on the right-hand side of the conservation laws, as well as in the moments for Ohm's law. Therefore, the fluid part can be advanced by using an appropriate numerical scheme for solving hyperbolic conservation laws. On the other hand, the kinetic species may be solved by the standard Particle-in-Cell (PIC) scheme. The feedback from the fluid onto the kinetic populations should appear only through the electromagnetic field.

10.6.1 Kinetic Part

The PIC scheme implemented for the kinetic part is fairly standard. We use the second-order shape function $S(\mathbf{x})$ for calculating the electromagnetic field on the particle positions from the field defined on the grid points. The particles' equation of motions is then pushed using the standard Buneman-Boris method.

The velocity moments are accumulated on the grid points using the same shape function. The zeroth-, first-, and second-order velocity moments for the particle species s denoted as I_s^0, I_s^1, I_s^2 at the position \mathbf{r} defined by

$$I_s^0(\mathbf{r}) = \sum_j m_s S(\mathbf{r} - \mathbf{x}_{s,j}), \quad (10.41)$$

$$I_s^1(\mathbf{r}) = \sum_j m_s \mathbf{v}_{s,j} S(\mathbf{r} - \mathbf{x}_{s,j}) \quad (10.42)$$

$$I_s^2(\mathbf{r}) = \sum_j m_s \mathbf{v}_{s,j} \mathbf{v}_{s,j} S(\mathbf{r} - \mathbf{x}_{s,j}) \quad (10.43)$$

are collected. Here, $(\mathbf{v}_{s,j}, \mathbf{x}_{s,j})$ represents the phase-space coordinate of the j -th particle of the species s , and the summation is taken over all particles. Notice that we have to calculate second-order moment \mathbf{I}_s^2 , which is not needed in a standard hybrid code. Considering symmetry of \mathbf{I}_s^2 , ten moments for each species must be stored in memory.

Once the moment quantities are given, the charge and current densities ϱ_k , \mathbf{J}_k as well as the moment quantities Λ_k , $\boldsymbol{\Gamma}_k$, $\boldsymbol{\Pi}_k$ may be obtained by:

$$\varrho_k = \sum_{s=kinetic} \frac{q_s}{m_s} I_s^0, \quad (10.44)$$

$$\mathbf{J}_k = \sum_{s=kinetic} \frac{q_s}{m_s} \mathbf{I}_s^1, \quad (10.45)$$

$$\Lambda_k = \sum_{s=kinetic} \frac{4\pi q_s^2}{m_s^2} I_s^0, \quad (10.46)$$

$$\boldsymbol{\Gamma}_k = \sum_{s=kinetic} \frac{4\pi q_s^2}{m_s^2} \mathbf{I}_s^1, \quad (10.47)$$

$$\boldsymbol{\Pi}_k = \sum_{s=kinetic} \frac{4\pi q_s}{m_s} \mathbf{I}_s^2. \quad (10.48)$$

Typically, a three-point binomial filter is applied to all the moment quantities to eliminate artificial noise before calculating these quantities.

10.6.2 Fluid Part

For solving the conservation laws and the magnetic induction equation, we employ the HLL (Harten-Lax-van Leer) approximate Riemann solver supplemented by the UCT (Upwind Constrained Transport) scheme. The resulting scheme, called the HLL-UCT [11, 12], ingeniously combines the one-dimensional Riemann solver (or the HLL solver) and the Constrained Transport (CT) scheme that guarantees the divergence-free condition for the magnetic field ($\nabla \cdot \mathbf{B} = 0$). The only spectral information required for the HLL-UCT scheme is the maximum phase speed. This is in contrast to more advanced multistate Riemann solvers such as the Harten-Lax-van Leer discontinuity (HLLD) scheme [13], which requires the eigenmode decomposition. We believe that the HLL scheme is better suited in practice for the present model because it will be extremely difficult or even impossible to perform the eigenmode decomposition.

For the QNTF equations (in the absence of the kinetic populations), the maximum phase speed in the plasma rest frame may be obtained by solving the dispersion relation (10.35). This choice has been successful in capturing

complicated multidimensional discontinuities and dispersive waves at the same time.

However, the situation is a little more involved in the presence of kinetic populations. More specifically, we find that the numerical dissipation is too excessive in such a case. This may be understood as due to artificial noise inherent in the PIC scheme. The discrete particle noise is primarily due to spontaneous emissions of plasma waves (or thermal fluctuations) generated by particle free streaming. These fluctuations may be reabsorbed by the particles, and ideally, the emission and absorption should balance with each other. However, since the numerical dissipation tends to damp out the fluctuations, the reabsorption by the particles will be underestimated. In other words, particles are artificially cooled down because they transfer the kinetic energy to the fluctuations (or waves), which in turn dissipates numerically.

To avoid the numerical cooling, we have to reduce the numerical dissipation in the presence of kinetic populations. We note that the amount of numerical dissipation in the HLL scheme is controlled by the maximum phase speed: The larger the maximum phase speed, the stronger the numerical dissipation. We thus choose the Alfvén speed V_A as the maximum phase speed instead of the solution obtained from the full dispersion relation (10.35), which, in general, gives a larger phase speed. We should note that this does not mean to be optimum but rather a compromise that works reasonably well.

10.6.3 Ohm's Law

Once the moments of both the fluid and kinetic populations are given, we can determine the electric field using the generalized Ohm's law. However, it is an implicit equation for the electric field as the left-hand side of Eq. (10.26) involves a spatial derivative. When discretized on the mesh, the equation becomes a linear matrix equation, which may be inverted by using either an iterative or a direct method. We adopt an approximation $\nabla \times \nabla \times \mathbf{E} \approx -\nabla^2 \mathbf{E}$ and use the second-order finite difference for the discretization of the Laplacian operator. (Note that we have assumed $\nabla \cdot \mathbf{E} = 0$, which is consistent with the quasi-neutrality assumption but is not numerically guaranteed.) This decouples the three vector components, and the left-hand side of Eq. (10.26) may be discretized in 3D as

$$\begin{aligned} (\Lambda \mathcal{E} - c^2 \nabla^2 \mathcal{E})_{i,j,k} &\approx (\Lambda_{i,j,k} + 2(\epsilon_x + \epsilon_y + \epsilon_z)) \mathcal{E}_{i,j,k} \\ &- [\epsilon_x (\mathcal{E}_{i-1,j,k} + \mathcal{E}_{i+1,j,k}) + \epsilon_y (\mathcal{E}_{i,j-1,k} + \mathcal{E}_{i,j+1,k}) + \epsilon_z (\mathcal{E}_{i,j,k-1} + \mathcal{E}_{i,j,k+1})], \end{aligned} \quad (10.49)$$

where $\mathcal{E} = E_x, E_y, E_z$ represents the three components of the electric field vector. Here, i, j, k indicates discretization indices in x, y, z directions, and $\epsilon_x = c^2/\Delta x^2, \epsilon_y = c^2/\Delta y^2, \epsilon_z = c^2/\Delta z^2$ with $\Delta x, \Delta y, \Delta z$ are grid sizes in each

direction. Likewise, the second-order central difference is used to evaluate spatial derivatives in the right-hand side, which gives the source vector to the matrix equation.

The resulting sparse matrix for each component can easily be inverted by an iterative method. We use the symmetric Gauss-Seidel method, which is quite easy to implement. Although the convergence of the method is not so fast, it is sufficient for a diagonally dominant matrix. It is easy to confirm that the matrix is diagonally dominant as long as the grid size is sufficiently larger than the electron inertial length. For higher resolutions or equivalently lower density regions, the equation approaches Laplace's equation (10.40) in a vacuum and the diagonal dominant property will be lost. More advanced methods such as the multigrid would be desired in these situations for better computational efficiency.

10.6.4 Time Integration

The time integration of the total system may be performed either by a Runge-Kutta or a predictor-corrector-type scheme. Typically, we employ the third-order TVD-RK (Total Variation Diminishing Runge-Kutta) scheme [14] for the QNTF equations, whereas the predictor-predictor-corrector scheme [15] is used if there exist a kinetic species to reduce the number of computationally expensive particle pushes. We note that methods specifically designed for the conventional hybrid codes, such as the CAM-CL (Current Advance Method and Cyclic Leapfrog) scheme [16], cannot be used here because of the different forms of the generalized Ohm's law.

10.7 Numerical Examples

10.7.1 Quasi-Neutral Two-Fluid Case

The first example is the well-known Brio-Wu shock tube problem [17] simulated with the QNTF model without kinetic populations. The purpose of this example is to demonstrate that the same code can consistently reproduce both the ideal MHD and dispersive regimes (e.g., [18, 19]) depending on the numerical resolution.

The simulation setup is identical to that presented in [1]. The MHD quantities are initialized in the same way as the ideal MHD with the total mass density and pressure defined as $\rho \equiv \rho_i + \rho_e$ and $p \equiv p_i + p_e$. For comparison with the self-similar MHD result, the box size is normalized to unity. A finite q_i/m_i then defines the ion inertial length and the ion cyclotron frequency. We use an ion-to-electron mass ratio of $m_i/m_e = 100$, and q_i/m_i is chosen such that the ion inertial length is given by $V_A/\Omega_{ci} = 10^{-3}/\sqrt{\rho}$. The solution should be closer to the ideal MHD result at a low resolution. Dispersive wave characteristics arising from the two-fluid effect will appear as increasing the numerical resolution. For this problem, we fix the ion-to-electron temperature ratio $\tau = T_i/T_e = 1$ so that the dissipated energy is always equally distributed to ions and electrons.

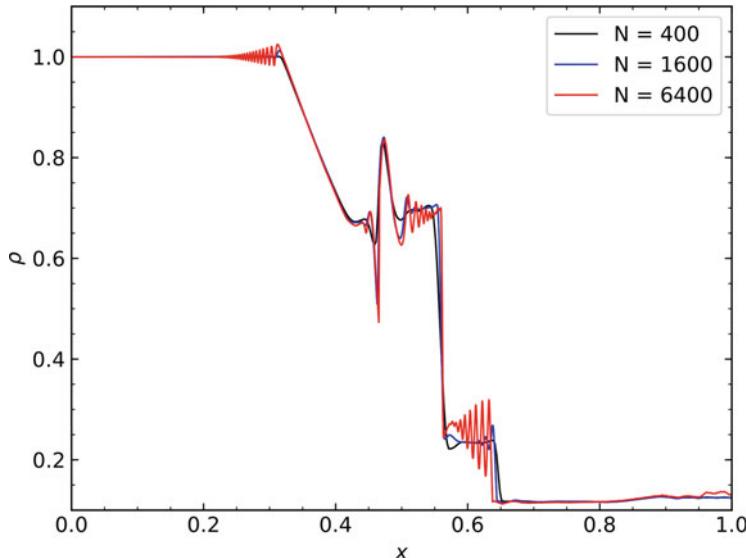


Fig. 10.2 Results obtained for the Brio-Wu shock tube problem. The mass density profiles at $t = 0.1$ for three different resolutions $N = 400, 1600, 6400$ are shown with black, blue, and red lines

Figure 10.2 shows the total mass density ρ at $t = 0.1$ for three different resolutions: $N = 400, 1600, 6400$ with N being the number of grid (or $\Delta x = 2.5 \times 10^{-3}, 6.25 \times 10^{-4}, 1.5625 \times 10^{-4}$ where $\Delta x = 1/N$.) Since the grid size at the lowest resolution was larger than the ion inertial length, dispersive characteristics were not clearly visible, and the result is almost the same as the ideal MHD. For higher resolutions, dispersive wave train structures were attached to the MHD discontinuities and rarefaction waves. These short wavelength fluctuations were also seen in earlier works that employed a full two-fluid model without the quasi-neutrality assumption (e.g., [20]).

This result demonstrates that the simulation code reduces to the ideal MHD whenever the grid size is larger than the ion inertial length. We should emphasize that numerical stability, in this case, is simply determined by the CFL (Courant-Friedrichs-Lowy) condition defined with respect to the phase velocities of MHD waves. There are no additional restrictions on the time step and the grid size that are specific to the QNTF model.

10.7.2 Hybrid and EP-MHD Hybrid Case

The next example demonstrates the capability of using the same code as a standard hybrid code, as well as an EP-MHD hybrid code. We can easily set up a standard hybrid simulation by putting $q_i/m_i = 0$ for the fluid ions so that the charge density

contribution becomes negligible. For an EP-MHD hybrid simulation, both the fluid and kinetic populations have finite densities.

As an example, we consider an instability driven by a small fraction of ion beam streaming along the ambient magnetic field in 1D [21]. This is one of the standard test problems for a hybrid code. In particular, we consider a regime in which the instability is driven by the cyclotron resonance of the beam population. If the beam drift velocity is sufficiently larger than the Alfvén speed, the resonance occurs at a wavelength much longer than the ion inertial length. This suggests that the background ions can be well described by the fluid approximation. Therefore, for an EP-MHD hybrid simulation, the background ions are approximated as a fluid whereas the fully kinetic effect is retained for the beam population.

In the following, the results of 1D simulations along x direction with the background magnetic field $\mathbf{B} = B_0 \mathbf{e}_x$ are presented. The ions consist of the core (background) and beam components with a relative streaming in the x direction. The normalized beam density is $n_b/n_0 = 0.02$ (where n_0 is the electron density), and the drift velocities are $V_b/V_A = 10$ and $V_c/V_A = -0.2$ for the beam and core, respectively. The simulations are performed in the rest frame of the electron fluid.

For the standard hybrid run, both distributions are represented by particles with the same charge-to-mass ratio q_s/m_s and the density of the fluid ions is ignored ($q_i/m_i = 0$). For the EP-MHD hybrid run, the core population is approximated by a fluid. The velocity distribution for each component is initialized by the Maxwellian distribution function in its rest frame. The thermal velocities are given by $v_{\text{th},b}/V_A = v_{\text{th},c}/V_A = 1/\sqrt{2}$. The electron plasma beta is $\beta_e = 1.0$. The other simulation parameters are $\Omega_{ci}\Delta t = 10^{-2}$, $\Delta x/c/\omega_{pi} = 0.25$, $N_x = 1024$, $N_{PPC} = 512$ (number of particles per cell—PPC—of each species). For the EP-MHD simulation, the core ion is replaced with the fluid, but otherwise, the setup is identical. In both simulations, we use a mass ratio of $m_i/m_e = 100$ and the polytropic equation of state for electrons $\rho_e \propto p_e^\gamma$ with $\gamma = 5/3$.

Both runs reproduced the linear growth of the instability quite well with only a slight delay of the instability onset in the EP-MHD run, which is probably due to lower artificial noise. Therefore, we proceed to compare results during the nonlinear stage of the instability. As a result of instability development, the parallel kinetic energy of the beam will be transferred to the wave and perpendicular energies. This is clearly illustrated in Fig. 10.3 showing the time history of parallel and perpendicular kinetic energies for the beam component. We see very good agreement between the two runs. Figure 10.4 compares the two snapshots at around the saturation. A nearly monochromatic electromagnetic wave is clearly seen in B_y (top). The core ions are nonresonant with respect to the wave, and thus the amplitude of perturbation is small (middle) in comparison with the beam component (bottom). As expected, the fluid approximation for the core component did not lead to a significant difference between the two runs.

We have seen that the EP-MHD run reproduced the result obtained by the standard hybrid quite well. The crucial point is that both simulations are run with the same code. With the appropriate setup, it runs as a standard hybrid or an EP-MHD hybrid code.

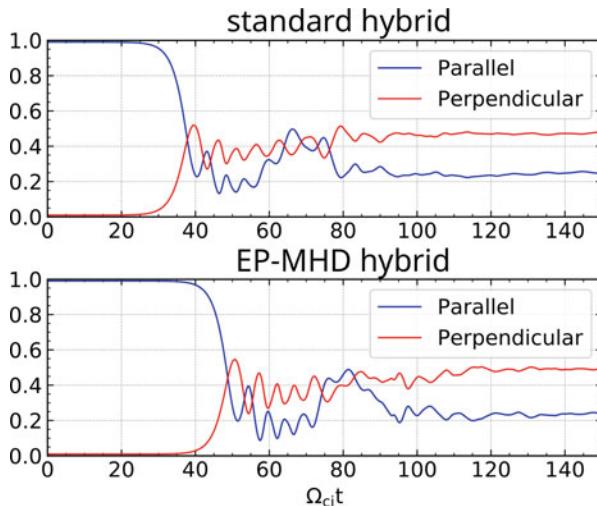


Fig. 10.3 Comparison of energy histories for the beam component obtained with standard hybrid (top) and EP-MHD hybrid (bottom) simulations

10.7.3 Fully Kinetic Energetic Electrons

Since the generalized Ohm's law is exact, the charge-to-mass ratio q_s/m_s is an arbitrary parameter. Therefore, the code correctly takes into account the kinetic effect even if the kinetic species is an energetic electron population. We should note that, however, the assumption of quasi-neutrality implies that the thermal electrons cannot be a kinetic species.

We consider an instability driven by a perpendicular temperature anisotropy $T_{\perp}/T_{\parallel} > 1$ of the energetic electron population. The anisotropy destabilizes whistler waves propagating parallel to the ambient magnetic field via the cyclotron resonance. Therefore, this instability may be reproduced only when the fully kinetic effect is taken into account.

To concentrate only on the kinetic effect associated with the energetic electrons, the background ions and electrons are approximated as fluids and only the energetic electrons with $m_i/m_e = 100$ are solved as a kinetic species. The energetic electron density is chosen as $n_{EP}/n_0 = 0.01$ where n_0 is the total density. The background ion and electron densities are given by $n_i = n_0 + n_{EP}$ and $n_e = n_0 - n_{EP}$, respectively. The initial temperature anisotropy is set to $T_{EP,\perp}/T_{EP,\parallel} = 9$ with a parallel beta of $\beta_{EP,\parallel} = 0.1$. The fluid ion and electron beta are as follows: $\beta_i = 0.1$, $\beta_e = \beta_i(1 - n_k/n_0) = 0.099$. The simulation is performed in a 1D simulation box along x direction with the following parameters: $\Omega_{ci}\Delta t = 5 \times 10^{-5}$, $\Delta x/c/\omega_{pi} = 2.5 \times 10^{-2}$, $N_x = 512$, $N_{PPC} = 1024$. We choose the time step to be small enough to resolve the electron cyclotron motion.

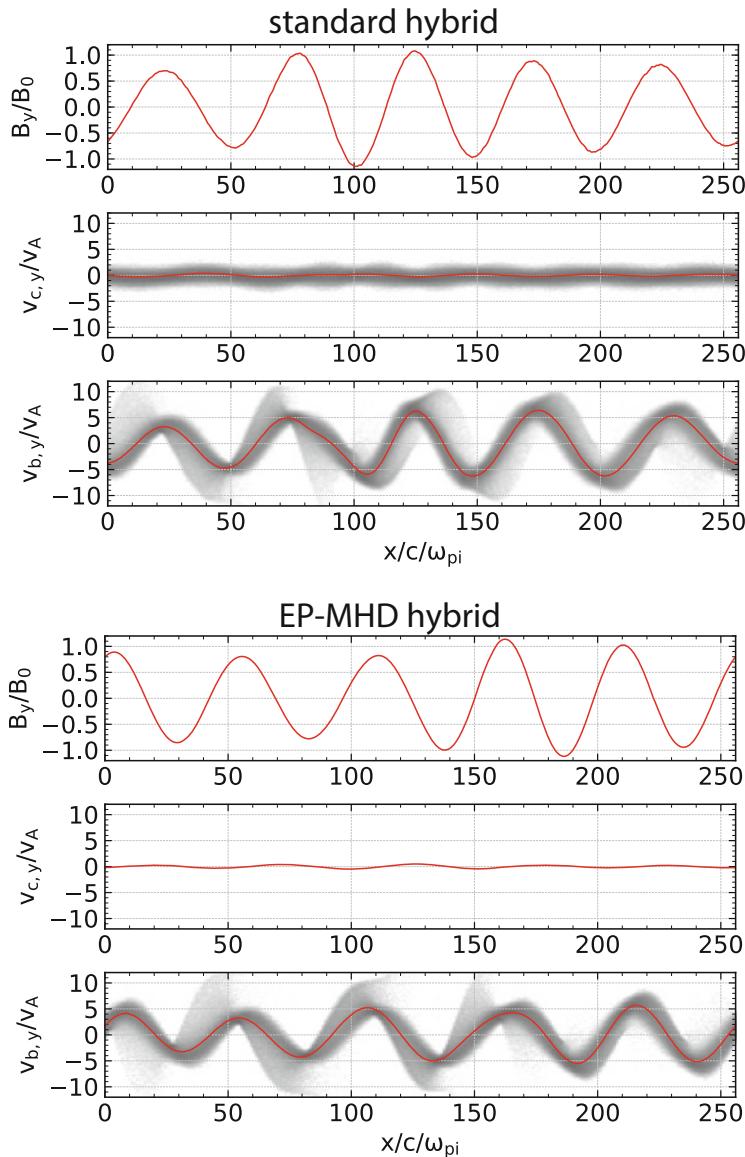


Fig. 10.4 Comparison of snapshots at around the saturation obtained by standard hybrid (top) and EP-MHD hybrid (bottom) simulations. The magnetic field B_y , core ion velocity $v_{c,y}$, and beam ion velocity $v_{b,y}$ are shown for each simulation. For kinetic species (core and beam for the standard hybrid, beam for the EP-MHD simulations), phase space plots are shown in gray scale

The early development of the simulation is well consistent with the prediction by the linear theory. Therefore, we again focus on nonlinear development. The time history of normalized wave power and temperature anisotropy is shown in Fig. 10.5. The wave growth is accompanied by the reduction of the anisotropy. It is important to point out that this self-consistent development can be observed only when the exact form of Ohm's law is utilized. For the purpose of comparison, the result without the off-diagonal terms of the $\boldsymbol{\Pi}$ tensor in the generalized Ohm's law is also shown with the dashed line. In this artificial setup, no wave growth or reduction of anisotropy is observed. This clearly indicates that the use of the exact Ohm's law is essential for fully kinetic treatment.

In the standard theoretical understanding, the reduction of anisotropy should be associated with pitch-angle scattering of electrons in resonance with the whistler waves. To confirm this, the deviation of the velocity distribution function from the initial condition $\delta f(v_{\parallel}, v_{\perp}, t) = f(v_{\parallel}, v_{\perp}, t) - f(v_{\parallel}, v_{\perp}, t = 0)$ is shown in Fig. 10.6. The distribution function is normalized such that the integral over velocity space becomes unity. From the linear theory, we may estimate the cyclotron resonance velocity for the wave of maximum growth rate to be $v_{\text{res}}/V_A \sim 26$. The scattering modifies the distribution function such that $\delta f < 0$ for $v_{\parallel} \lesssim v_{\text{res}}$ and $\delta f > 0$ for $v_{\parallel} \gtrsim v_{\text{res}}$. The pitch angles of the particles become more field-aligned as a result, leading to the net reduction of the temperature anisotropy. This

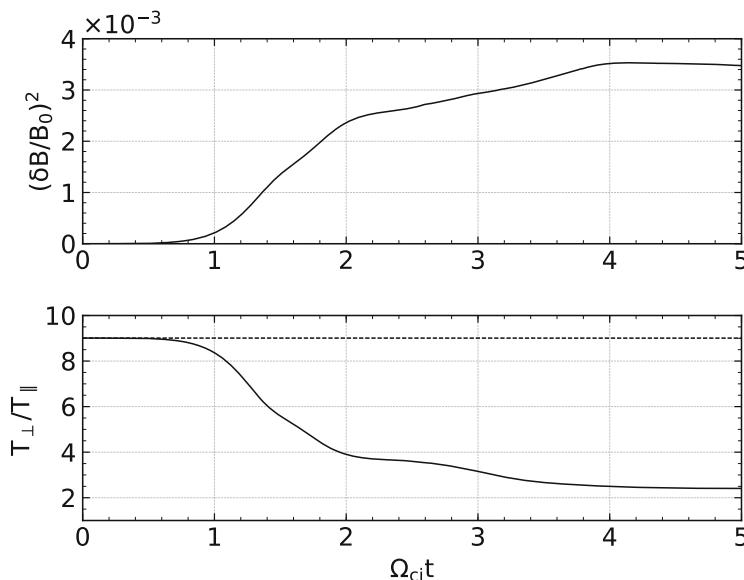


Fig. 10.5 Time histories of normalized fluctuating magnetic field energy density and temperature anisotropy. The dashed line represents the result obtained without the off-diagonal components of $\boldsymbol{\Pi}$ tensor in the generalized Ohm's law

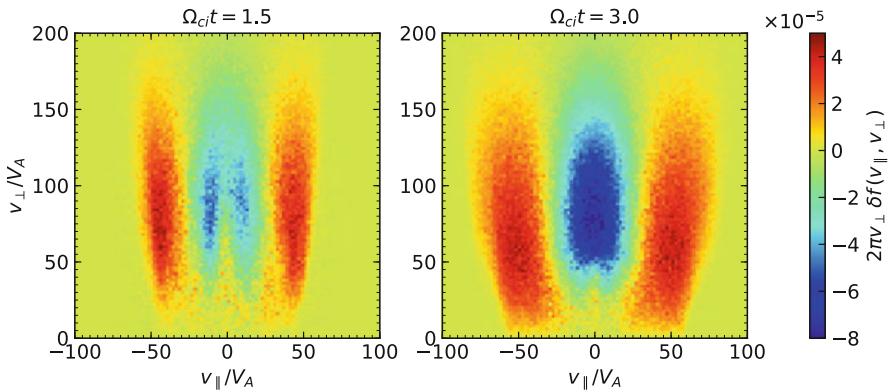


Fig. 10.6 Deviation of velocity distribution function from the initial condition $2\pi v_\perp \delta f(v_\perp, v_\parallel)$. Two snapshots at $\Omega_{ci}t = 1.5, 3.0$ are shown

is exactly what is seen in Fig. 10.6, confirming that the kinetic effect associated with the energetic electrons is correctly included in the code.

10.8 Summary and Outlook

We have presented the GQNH model as a generalized version of the hybrid simulation model under the assumption of quasi-neutrality. The quasi-neutrality eliminates high-frequency Langmuir waves and free space electromagnetic waves from the system and allows us to concentrate on low-frequency phenomena. Since the exact form of the generalized Ohm's law is utilized, it correctly describes both ion and electron inertial effects, and the highest physical frequency is the electron cyclotron frequency $\omega \sim \Omega_{ce}$. It has been shown that the conventional plasma simulation models (under the assumption of quasi-neutrality) are the special cases of the GQNH model. Therefore, a single numerical simulation code that solves the GQNH equations can serve as a useful alternative to conventional codes.

Perhaps, what is more important is that the model may provide a framework for multi-scale and/or multi-physics simulations for space and astrophysical plasma modeling in general. The model automatically reproduces dispersive waves arising from finite ion and electron inertial effects whenever needed, whereas it is essentially equivalent to the ideal MHD in the other regions. Furthermore, we can seamlessly deal with kinetic regions (where the density is dominated by kinetic species) and fluid regions (where the density is dominated by fluid species) and their spatial boundaries without any gaps. This is indeed a desirable feature in recent and future space plasma modeling.

One of the limitations of the model is that it, by assumption, has to separate fluid and kinetic populations in the first place, and there is no self-consistent injection from the cold fluid to the energetic particles (unless an ad hoc injection recipe

is used). Nevertheless, it is still possible to consider the propagation of energetic particles accelerated in a fully kinetic region into a region where the dynamics is, in the absence of energetic particles, essentially described by MHD. In such a case, one may investigate the coupling between the MHD fluid and the self-consistently generated energetic particles with this model.

One of the examples is the problem of magnetic reconnection, which is a key process in magnetospheric physics that controls even the global dynamics. The kinetic effects are thought to be crucial at a microscopic diffusion region of magnetic reconnection. One of the popular approaches is to embed a small domain in which the fully kinetic PIC code applies in a much larger domain that evolves according to the ideal MHD. In principle, this should give a reasonable result if the boundary between the two regions is appropriately connected. However, this is, in general, very difficult to achieve. For the reconnection problem, energetic particles generated around the diffusion region may propagate along the field lines. They can carry energy and momentum to large distances and thus provide a heat flux in the macroscopic sense. Application of this approach should thus be limited to problems in which propagation of beyond-MHD effects are negligible at a certain distance from the key region.

Although a fully kinetic PIC domain is beyond the limit of applicability, hybrid and MHD domains can be smoothly connected (there is no boundary indeed) with the present model. It is possible to take into account the self-consistent dynamical effects of energetic particles escaping from the kinetic domain even in the much larger MHD domain. In short, with the GQNH model, we may be able to model multi-scale and multi-physics phenomena in a physically more consistent fashion. For better modeling of microphysics associated with the electron dynamics, it is also important to adopt more sophisticated equation of states (including pressure anisotropy or non-gyrotropy), which indeed changes the dynamics of magnetic reconnection.

Acknowledgments The work of T.A. has been partially supported by JSPS KAKENHI Grant Numbers 25800101, 16H01170, 17H02966, and 17H06140.

References

1. T. Amano, J. Comput. Phys. **299**, 863 (2015). <https://doi.org/10.1016/j.jcp.2015.07.035>. <http://www.sciencedirect.com/science/article/pii/S0021999115004805>
2. T. Amano, J. Comput. Phys. **366**, 366 (2018). <https://doi.org/10.1016/j.jcp.2018.04.020>
3. D.E. Shumaker, D.V. Anderson, G.F. Simonson, Unknow (1992). <https://ui.adsabs.harvard.edu/abs/1992tds.rept.....S/abstract>
4. A.S. Lipatov, *The Hybrid Multiscale Simulation Technology: An Introduction with Application to Astro Physical and Laboratory Plasmas* (Springer, Berlin, 2002)
5. T. Amano, K. Higashimori, K. Shirakawa, J. Comput. Phys. **275**, 197 (2014). <https://doi.org/10.1016/j.jcp.2014.06.048>. <https://adsabs.harvard.edu/abs/2014JCoPh.275..197A>
6. D.S. Harned, J. Comput. Phys. **47**, 452 (1982). [https://doi.org/10.1016/0021-9991\(82\)90094-8](https://doi.org/10.1016/0021-9991(82)90094-8)

-
7. D.W. Hewett, J. Comput. Phys. **38**, 378 (1980). [https://doi.org/10.1016/0021-9991\(80\)90155-2](https://doi.org/10.1016/0021-9991(80)90155-2)
 8. M. Holmström, in *Numerical Modeling of Space Plasma Flows (ASTRONUM2012)*, *Astronomical Society of the Pacific Conference Series*, vol. 474, ed. by N.V. Pogorelov, E. Audit, G.P. Zank. Astronomical Society of the Pacific Conference Series, vol. 474 (2013), p. 202
 9. A.L. Zachary, B.I. Cohen, J. Comput. Phys. **66**(2), 469 (1986). [https://doi.org/10.1016/0021-9991\(86\)90076-8](https://doi.org/10.1016/0021-9991(86)90076-8). <http://adsabs.harvard.edu/abs/1986JCoPh..66..469Z>
 10. X.N. Bai, D. Caprioli, L. Sironi, A. Spitkovsky, Astrophys. J. **809**(1), 55 (2015). <https://doi.org/10.1088/0004-637X/809/1/55>. <http://adsabs.harvard.edu/abs/2015ApJ...809...55B>
 11. P. Londrillo, L. Del Zanna, J. Comput. Phys. **195**(1), 17 (2004). <https://doi.org/10.1016/j.jcp.2003.09.016>. <http://adsabs.harvard.edu/abs/2004JCoPh.195...17L>
 12. P. Londrillo, L. Del Zanna, J. Comput. Phys. **195**, 17 (2004). <https://doi.org/10.1016/j.jcp.2003.09.016>
 13. T. Miyoshi, K. Kusano, J. Comput. Phys. **208**, 315 (2005). <https://doi.org/10.1016/j.jcp.2005.02.017>
 14. C.W. Shu, S. Osher, J. Comput. Phys. **77**, 439 (1988). [https://doi.org/10.1016/0021-9991\(88\)90177-5](https://doi.org/10.1016/0021-9991(88)90177-5)
 15. M.W. Kunz, J.M. Stone, X.N. Bai, J. Comput. Phys. **259**, 154 (2014). <https://doi.org/10.1016/j.jcp.2013.11.035>. <http://adsabs.harvard.edu/abs/2014JCoPh.259..154K>
 16. A.P. Matthews, J. Comput. Phys. **112**, 102 (1994). <https://doi.org/10.1006/jcph.1994.1084>
 17. M. Brio, C.C. Wu, J. Comput. Phys. **75**, 400 (1988). [https://doi.org/10.1016/0021-9991\(88\)90120-9](https://doi.org/10.1016/0021-9991(88)90120-9)
 18. A. Hakim, J. Loverich, U. Shumlak, J. Comput. Phys. **219**, 418 (2006). <https://doi.org/10.1016/j.jcp.2006.03.036>
 19. H. Kumar, S. Mishra, J. Sci. Comput. **52**(2), 401 (2012). <https://doi.org/10.1007/s10915-011-9554-7>
 20. A. Hakim, J. Loverich, U. Shumlak, J. Comput. Phys. **219**(1), 418 (2006). <https://doi.org/10.1016/j.jcp.2006.03.036>. <http://adsabs.harvard.edu/abs/2006JCoPh.219..418H>
 21. D. Winske, M.M. Leroy, J. Geophys. Res. **89**, 2673 (1984). <https://doi.org/10.1029/JA089iA05p02673>



Fully Kinetic (Particle-in-Cell) Simulation of 11 Astrophysical Plasmas

Masahiro Hoshino

Abstract

The methodology of particle-in-cell (PIC) simulations is useful for studying not only space plasma in the Earth's neighborhood but also astrophysical plasmas in the universe. The typical plasma parameters of the high-energy phenomena observed in astrophysical settings are vastly different from those in the solar system, but many fundamental plasma processes can be shared to obtain a better understanding of the plasma universe. This is because of the universal validity of the Maxwell-Vlasov system in collisionless plasmas. This chapter addresses several PIC simulation studies that focus mainly on plasma heating and particle acceleration seen in the astroplasma environment.

11.1 Introduction

Over the last few decades, magnetohydrodynamic (MHD) and particle-in-cell (PIC) simulation studies have been recognized as an important tool for investigating astroplasma phenomena such as supernova shocks with high Mach numbers [1, 2], synchrotron radiation from pulsar wind nebula [3, 4], gamma-ray burst (GRB) afterglow from radio to X-ray frequencies [5, 6], particle acceleration in relativistic shocks in active galactic nucleus (AGN) jets [7], and fast radio bursts (FRBs) with millisecond durations [8]. As full PIC simulations resolve the Debye length and gyro-motion, it cannot necessarily deal with large-scale dynamics in the universe, whose scale is many orders of magnitude larger than the system size in PIC simulations. However, PIC simulations can address many important problems

M. Hoshino (✉)
University of Tokyo, Tokyo, Japan
e-mail: hoshino@eps.s.u-tokyo.ac.jp

beyond the MHD description successfully, such as the origin and generation of high-energy particles, mechanism of energy dissipation, and electromagnetic radiation. The kinetic transport processes involved in viscosity, resistivity, and thermal conduction, which play an important role in macroscale physics, can be self-consistently described in PIC simulations.

Typical parameters discussed in astrophysics are usually vastly different from those in space plasma. The most conspicuous example may be the relativistic effects that often play an important role in plasma dynamics in astrophysics. In space plasma, relativistic effects are sometimes partially discussed in the physics of energetic electron accelerations in the radiation belt, the magnetotail reconnection, and so on. However, relativistic bulk flow with a large Lorentz factor and/or relativistically hot plasma whose temperature is larger than the rest mass energy are not necessarily present. However, owing to the universal validity of plasma physics, we share many fundamental plasma problems of space plasma physics and astroplasma physics to obtain a better and clearer understanding of the plasma universe [9]. In this chapter, we briefly demonstrate some astrophysical research problems using PIC simulations as an interesting interdisciplinary research field. In addition, we briefly mention some numerical techniques that are necessary for studying astroplasma phenomena.

11.2 Astrophysical Shock Waves

11.2.1 Non-relativistic, High-Mach-Number Shocks

Supernova shocks are widely believed to be the source of cosmic rays with energies up to approximately $10^{15.5}$ eV. X-ray synchrotron emission and TeV-scale gamma rays are also detected in supernova remnants (SNRs). During a supernova explosion, the ejector expanding into the interstellar medium (ISM) at speeds of a few thousand km/s drives a high-Mach-number shock with $M > 100$. The diffusive shock acceleration (DSA)/Fermi acceleration process is widely believed to operate in these high-Mach-number shocks, leading to the generation of the observed cosmic rays and synchrotron emissions [10, 11]. On the other hand, the observed energy spectrum of the cosmic rays can be approximated by $N(\varepsilon) \propto \varepsilon^{-s}$ with $s \sim 2.7$; the theoretical prediction of the energy spectrum is given by $s = (r+2)/(r-1)$, where r is the plasma density compression ratio. For $M \gg 1$, $r = 4$ and the power-law index $s = 2$, which is close to the observation. The difference between the indices 2 and 2.7 may be attributed to the loss process during the propagation of cosmic rays in the ISM. The DSA model is widely accepted because the power-law index is independent of detailed physical quantities such as plasma temperature, magnitude of the magnetic field, etc.

Although DSA theory can explain the observed cosmic ray spectrum, many issues remain unsolved. One such issue of the shock acceleration scenario is the pre-acceleration process of the thermal plasma that is necessary before the DSA process can operate. Only the pre-accelerated particles can cross the shock front

back and forth, upstream and downstream of the shock. Other thermal particles are only transported downstream in association with the bulk plasma flow and remain thermal plasma. The pre-acceleration process, which is called the shock-injection problem [12, 13], remains an enigma that needs to be understood from the viewpoint of kinetic plasma.

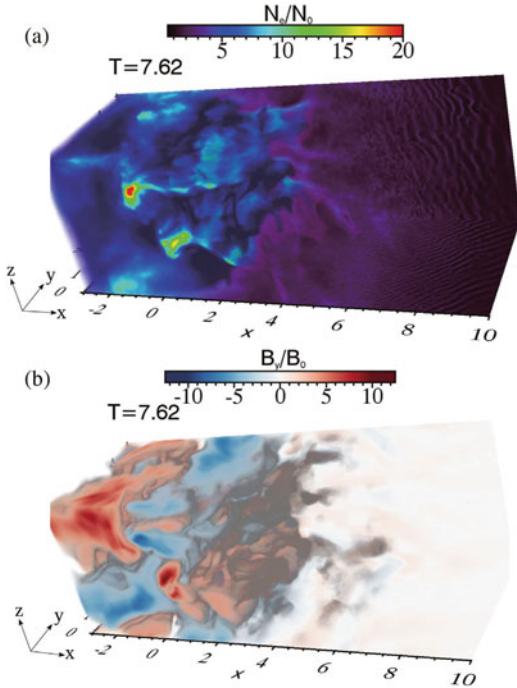
To understand the shock-injection physics of the shock front, where highly nonlinear waves are generated, the differences between ion- and electron-injection need to be considered. The most important difference is the effect of the shock potential induced by an ambipolar electric field directed toward the shock upstream [14, 15]. Owing to the inertia difference of the incoming ions and electrons convected from upstream, the ions can penetrate more deeply into the shock downstream than the electrons, generating a positive shock potential. Therefore, the electron reflection from the shock front or electron leakage from the shock downstream needs to overcome the shock potential gap. As the shock potential is roughly estimated to be several ten percent of the incoming ion bulk flow energy [16], the electron-injection process requires stronger preheating/acceleration compared to the ion-injection process.

While the injection problem is one of the most important subjects studied in PIC simulation, another purpose of PIC simulations for shock physics is to discover new heating and acceleration processes caused by collective plasma phenomena. This is because the collisionless shock wave is a test bed of rich nonlinear phenomena, where many processes are yet to be understood.

To create a collisionless shock in a PIC simulation, supersonic uniform plasmas are continuously injected from one upstream boundary on either the right- or the left-hand side of the simulation domain, and they are transported toward the other end as the downstream boundary, where all particles and waves are reflected [17]. The counter-streaming plasma formed near the downstream boundary is usually unstable for two-stream instabilities, and the thermalized and compressed downstream plasma can be quickly generated. Then, the downstream plasma launches a compressional shock wave, which propagates toward the upstream region. In addition to the above method to initiate a shock wave, other methods such as the counter-streaming colliding flow generation and the magnetic-piston launch can be utilized [9].

So far, electron-injection and succeeding particle acceleration processes have been studied by many researchers using PIC simulations, for example, [13, 18–23, 114]. Figure 11.1 shows an example that focuses on the generation of nonlinear waves and their interaction with electrons [24] and their application to a supernova shock whose Mach number is often greater than 100. The following parameters are adopted in the simulation: the ion-to-electron mass ratio is 64, upstream plasma flow is $v_1/c \sim 0.26$, sonic and Alfvén Mach numbers are $M_s = v_1/c_s \sim 22.8$ and $M_A = v_1/V_A \sim 20.8$, respectively, and the shock angle is $\Theta_{BN} = \cos^{-1}(\mathbf{B} \cdot \mathbf{n}/|B|) = 74.3^\circ$, where \mathbf{n} is the shock normal vector. Thus, the shock wave is categorized as a subluminal shock with $\tan^{-1}(c/v_1) > \Theta_{BN}$, where relativistic particles following the magnetic field can escape ahead of the shock.

Fig. 11.1 Snapshot of the shock-front structure of the (a) electron density and (b) magnetic field obtained using a three-dimensional (3D) high-Mach-number shock. The Alfvén Mach number and shock angle are $M_A = 20.8$ and $\Theta_{BN} = 74.3^\circ$, respectively. The shock front is situated around $x = 0$, and the supersonic magnetized flow is injected from the right-hand boundary. The spatial scale and time are normalized by the ion inertia length (v_A/Ω_{ci}) and the inverse of the ion gyro-frequency (Ω_{ci}^{-1}), respectively. The simulation domain size is approximately $55 \times 5 \times 5$, and only the shock-front region is shown. The figure was reproduced from [24] with the permission of APS publishing



The left-hand side region corresponds to the downstream region of the compressed shock, while the right-hand side is the upstream region, where a magnetized supersonic plasma approaches the shock front. The simulation was performed in the downstream frame. The time and space are normalized by the inverse of the gyro-frequency of the ion, Ω_{ci}^{-1} , and the ion inertia length, V_A/Ω_{ci} .

Two distinct density perturbations can be observed: a coherent wave train located near $x = 7\text{--}10$, with a short wavelength whose scale is lesser than the ion inertial length and a largely fluctuating density clump in front of the shock whose scale is of the order of the ion inertial length. The former density fluctuations are known to be associated with a series of large-amplitude bipolar electrostatic fields (E_{bp}) generated by the Buneman instability between the reflected and incoming electrons [18–20, 25]. The latter can be attributed to the ion Weibel instability that is excited by the effective temperature anisotropy between the reflected and incoming ions [26, 27]. These two instabilities require a high Mach number, such as that found in supernova shocks, and may not be detected in a low-Mach-number shock, such as in the Earth’s bow shock. The simple criterion of the large-amplitude electric field with $E_{bp} > B_0$ in CGS units via the Buneman instability may be given as $M_A > (m_i/m_e)^p$ with $p = 3/5 \sim 2/3$, where B_0 is the ambient upstream magnetic field [23, 27–29], while the criterion of the onset of the ion Weibel instability is $M_A \gg 10$ [30].

In the shock-front region of high-Mach-number quasi-perpendicular shocks, the electron energization owing to interactions with the Buneman instability is understood as the shock-surfing acceleration [14, 28, 31, 32]. While a bipolar electric field E_{bp} generated by the Buneman instability can contribute to trapping the electrons inside the bipolar electric field, the bipolar fields can almost remain in the shock transition region against the incoming upstream plasma because they are generated by releasing the free energy of the reflected ions moving upstream. Therefore, the trapped electron can gain energy from the motional electric field $E = -v_r \times B/c$, where v_r is the upstream flow speed in the frame moving with the bipolar electric field. In addition to the pre-acceleration in the Buneman wave/bipolar electric field zone generated by the Buneman instability, the ion Weibel wave zone plays an important role in boosting electron energization. The typical scale of the ion Weibel waves is the ion inertia length v_A/Ω_{ci} , and the electrons pre-accelerated by the shock-surfing acceleration can interact with the ion Weibel waves. In a two-dimensional (2D) PIC simulation with improved spatial resolution by Matsumoto et al. [27], it was noted that many magnetic islands were formed by magnetic reconnection in the current sheet with the antiparallel magnetic fields generated by the ion Weibel instability, and the electrons interacting with these magnetic islands can gain their energies through a stochastic acceleration process [33, 34].

It is interesting to note that large-amplitude coherent waves, such as the Buneman and ion Weibel waves, are generally embedded in broadband turbulent waves, and the energization by interactions with such coherent electromagnetic and electrostatic waves plays a significant role in the injection stage. For the surfing acceleration zone, the wavelength of the Buneman wave is quite small beyond the MHD scale, but the electric field E_{bp} may become 100 times larger than the shock motional electric field in a high-Mach-number shock. This suggests that multiscale structuring phenomena under the coupling of macroscale and microscale physics appear to control many important processes such as nonthermal particle acceleration and energy partition between ion and electron [35].

So far, PIC simulations of non-relativistic, quasi-perpendicular, high-Mach-number shocks have provided many insights into the dynamic shock structures [36], but the shock propagation speeds are chosen to be higher than 10% of the speed of light owing to the reduced mass ratio of the ion to the electron. Simulations with a realistic shock speed of a few % of the speed of light are necessary.

11.2.2 Relativistic Shocks with a Large Bulk Lorentz Factor

Cosmic rays with energies up to $10^{15.5}$ eV are believed to be generated by supernova shocks in our galaxy, whereas the more energetic cosmic rays are considered to arise from extragalactic sources. Relativistic shock is the most plausible candidate for very-high-energy cosmic rays. For example, an AGN with a super massive black hole can form relativistic well-collimated jets whose bulk Lorentz factor

Γ is approximately 10 and is believed to generate ultra-high-energy cosmic rays (UHECRs) with energies of the order of 10^{20} eV.

In addition to AGN jets, many relativistic plasma flows that emanate from the central astrophysical objects are known in the plasma universe. Pulsars are believed to have a relativistic plasma wind whose bulk Lorentz factor Γ is of the order of 10^6 – 10^7 [3, 4]. Gamma-ray bursts (GRBs) that emit intense electromagnetic radiation during star collapse are believed to form a collimated relativistic jet with $\Gamma \sim 10$ [5, 6]. The particle acceleration in these relativistic shocks has been investigated for the relativistic version of the diffusive shock acceleration in a test particle approximation, but the kinetic plasma behavior is yet to be solved. Here, we discuss some kinetic aspects observed in the relativistic shock.

Multiple simulation studies of relativistic shocks have been performed, for example, [37–44], and their methodology is almost identical to that of non-relativistic PIC simulations. In most simulations, relativistic plasma flow is injected from an upstream boundary, and the particles and waves are assumed to be reflected at the downstream boundary. Compared to non-relativistic shocks, a long computational CPU time is not necessarily required, because the phase speed of the magnetosonic wave is of the same order as the speed of light c and the relativistic plasma frequency and cyclotron frequency are decreased as the particle energy increases. Thus, we can follow nonlinear time evolution with a large computational time step Δt , which satisfies the computational Courant–Friedrichs–Lewy condition $c\Delta t/\Delta x < 1$.

Early relativistic PIC simulations investigated the quasi-perpendicular shock geometry, where the shock normal vector is perpendicular to the upstream magnetic field [37–39] because most astrophysical relativistic shocks after the Lorentz transformation to the rest frame of the shock may become quasi-perpendicular shocks. In the perpendicular geometry, the standard diffusive shock acceleration does not occur except in extremely strong turbulence. Attempts are made to determine whether any other acceleration process operates in the relativistic shocks. It is argued that a relativistic shock with the Lorentz factor of upstream flow $\Gamma \gg 1$ can emit large-amplitude electromagnetic precursor waves that propagate upstream at almost the speed of light. Their generation in the shock front by means of synchrotron maser instability is discussed [45].

After the discovery of the intense electromagnetic wave emission in the upstream shock, it was argued that the ponderomotive force of the electromagnetic waves causes the electrons to lag behind ions. The ponderomotive force f_{pond} is a kind of wave pressure under a coherent wave, which is given by $f_{\text{pond}} = -(e^2/4m\omega^2)\nabla E^2$, where m , E , and ω are the mass of the particle, amplitude of the electromagnetic wave, and the wave frequency, respectively [41, 46]. Owing to the different responses of ions and electrons, an electrostatic field is generated to maintain the charge neutrality of the plasma, and then electrons can be accelerated up to the ion kinetic energy [40, 41]. Furthermore, during the nonlinear stage of the interaction between the intense electromagnetic waves and plasma, Hoshino [41] discussed the possibility of nonthermal particles being generated in a process similar to wakefield acceleration (WFA) [47], in which the electromagnetic wave (photon) decays into a Langmuir wave (plasmon) via the parametric decay instability [48]. It is to be noted

that in laboratory laser-plasma experiments, a power-law energy spectrum with a spectral index of 2 can be obtained during the nonlinear stage after the collapse of the Langmuir wave in the WFA [49]. Relativistic shocks are suggested as promising candidates for UHECRs.

In the early one-dimensional shock simulations, the electromagnetic precursory wave is found to be generated by the synchrotron maser instability (SMI). However, the coherent emission process of the SMI might be disguised by other shock-front instabilities in a multidimensional system, and the emission may be turned off. With the advances in supercomputers, 2D and 3D simulation studies of relativistic shocks have become possible, and the precursor wave emission in 2D PIC simulations has been investigated by Iwamoto et al. [50, 51].

However, we need to pay special attention to the numerical Cherenkov instability (NCI) [52–57]. NCI develops quickly in a system with relativistic plasma flow, after which numerical plasma heating occurs. Shock simulations are usually performed in the shock downstream frame, and the upstream relativistic plasma flow is always subject to NCI.

NCI is caused by the interaction between the relativistic bulk flow and electromagnetic waves for both the physical waves within the Nyquist wavenumber ($k < \pi/\Delta$, where Δ is the grid size) and their aliasing waves. In a finite-difference scheme, the phase speed of the electromagnetic wave becomes lower than the speed of light for large wavenumbers. The entropy mode for a high-speed bulk flow then interacts with the numerically dispersive electromagnetic wave. Even when adopting a Fourier pseudo-spectral scheme without any numerical dispersion error, the aliasing waves of the bulk flow entropy mode resonate with the electromagnetic waves at $c^2(k_x^2 + k_y^2) = v_b^2(2\pi/\Delta - k_x)^2$, where the left-hand side term shows the electromagnetic wave and the right-hand side term represents the aliasing wave of the entropy wave. We have assumed that the plasma flow with v_b is parallel to the x axis. One can easily find that the resonance between two modes appears in the obliquely propagating waves. Therefore, NCI is an inevitable numerical instability in PIC simulations under finite-difference methods. To suppress NCI, several approaches are proposed [53–58] (see chapter 6 on particle-in-cell (PIC) kinetic codes).

The shock structure shown in Fig. 11.2 is obtained by using the method to substantially suppress NCI by choosing a certain ratio of the time step to the grid size [58]. The orientation of the upstream magnetic field is chosen to be in the in-plane y direction. The bulk Lorentz factor of the upstream flow and the magnetization parameter are $\Gamma_1 = 1/\sqrt{1 - v_1^2/c^2} = 40$ and $\sigma = B_1^2/4\pi N_1 \Gamma_1 m_{\pm} c^2 = 3 \times 10^{-3}$, respectively. Here, N_1 and m_{\pm} represent the upstream plasma density in the simulation frame and the mass of the pair plasma, respectively. A cold magnetized plasma is injected from the right-hand boundary, and all particles and waves are reflected back at the left-hand boundary. The downstream electron density N_2 is compressed to the Rankine–Hugoniot state with four times the upstream density N_1 in the shock downstream frame, and we can observe a large density fluctuation around the shock front. In the second panel of B_x , we can see a similar

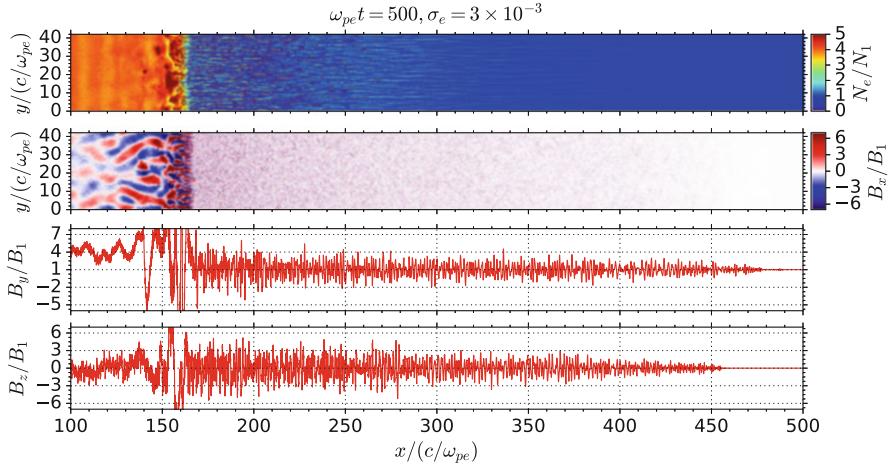


Fig. 11.2 Snapshot of a weakly magnetized relativistic shock structure in pair plasma. From top to bottom, the plasma density and the x component of the magnetic field in the $x - y$ plane, the one-dimensional slice of the y and z components of the magnetic field at $y/(c/\omega_{pe}) = 21$. The figure was reproduced from [59] and [24] with the permission of AAS publishing

fluctuation along the shock front, which is generated by the Weibel instability under the temperature anisotropy owing to the coexistence of the incoming and reflected particles [60]. The two lower panels depict plots of the one-dimensional cut of B_y and B_z . As we can see, the large-amplitude precursor waves for both the X-mode in B_y and O-mode in B_z are persistently observed, even if the Weibel instability occurs at the shock front. The large-amplitude precursor waves seen in the 2D pair plasma shocks are favorable for working the ponderomotive force, but the ion-electron plasma is necessary to initiate the longitudinal electric field/Langmuir wave. Iwamoto et al. [59] have recently demonstrated the existence of the WFA in a two-dimensional ion-electron shock and confirmed that the precursor wave amplitude can be significantly enhanced by a positive feedback process between the precursor waves and the pre-accelerated electron flow owing to WFA [41]. The SMI activity in the shock front increases with increasing electron acceleration efficiency in the precursor wave zone, while the electron acceleration can be enhanced by the SMI emission.

So far, precursor emission and particle acceleration have been introduced in the context of a possible acceleration process in quasi-perpendicular shocks. However, Metzger et al. [61] have recently proposed that the synchrotron maser instability (SMI) at the relativistic shock can explain the astrophysical phenomena of fast radio bursts (FRBs), which possess a large dispersion measure and transient radio pulse whose duration is less than a few milliseconds [8].

In the above, we have introduced some advances in the case of a quasi-perpendicular shock with a finite ambient background magnetic field. The formation of an unmagnetized shock without any ambient magnetic field has also been

investigated, as we believe that it can be applied, for example, to the GRB afterglow. The SMI with a finite magnetic field can provide energy dissipation in a magnetized shock, but the production of the entropy necessary to maintain an unmagnetized shock is not clearly understood. Spitkovsky [42] and Kato and Takabe [43] argued that the Weibel instability at the shock front suffices for the entropy production and shock formation.

We did not discuss diffusive shock acceleration, which is regarded as the standard theory of particle acceleration for both non-relativistic and relativistic regimes. Diffusive shock acceleration is the most successful model for explaining a universal power-law energy spectrum after the injection of supra-thermal particles into the Fermi acceleration process. So far, PIC simulation studies have focused more or less on the so-called wave-particle interaction to self-consistently understand the process by which the particle is scattered by waves in the shock system. On the other hand, the simulation study of diffusive shock acceleration requires time-consuming calculations because the diffusion time scale for particles moving back and forth across the shock front is many orders of magnitude slower than the cyclotron period of the particle.

It is noteworthy, however, that the fundamental process of diffusive shock acceleration in which supra-thermal particles undergo back-and-forth motion across the shock front has been demonstrated in hybrid/PIC simulations [42, 44, 62].

11.3 Magnetic Reconnection in Astroplasma Settings

11.3.1 Relativistic Magnetic Reconnection with Strong Magnetic Fields

In the astrophysical context, not only the particle acceleration of shock waves by releasing the supersonic bulk flow energy but also the rapid energy release of the magnetic field and its particle acceleration are regarded as important processes, for example, [63]. Specifically, particle acceleration and rapid electromagnetic energy dissipation in a relativistic magnetic reconnection have received attention owing to the great variety of astrophysical objects storing huge magnetic energy without the bulk flow energy. This is because reconnection can explain a harder energy spectrum than that modeled by the standard diffusive shock acceleration mechanism, for example [64, 65]. In the context of space, solar, and laboratory plasma physics, non-relativistic magnetic reconnection has been studied extensively, and it is known that large-scale magnetic field energy can be quickly released by changing the magnetic field topology. Relativistic reconnection is expected to release electromagnetic energy as well, and the efficiency of energy release would be much better compared to non-relativistic reconnection. The pulsar magnetosphere, magnetar, and AGN are examples of such high-energy astrophysical sites where relativistic reconnection with a strong magnetic field occurs.

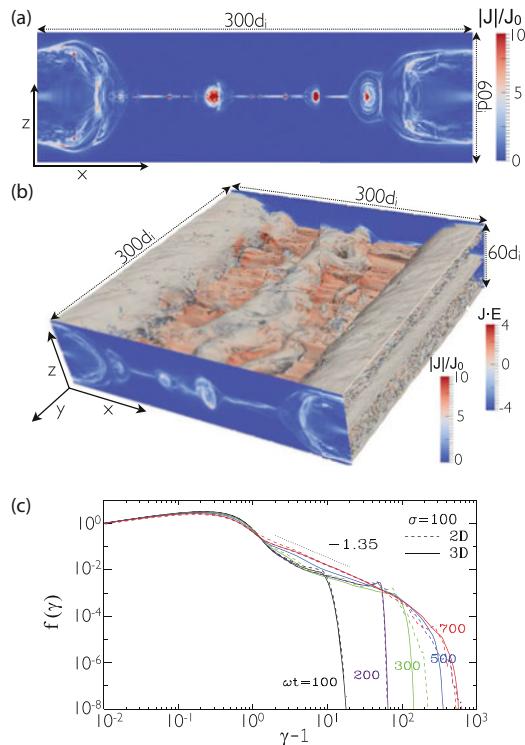
The relativistic reconnection process may be parameterized by the magnetization parameter $\sigma = B^2/4\pi Nmc^2$, which is the ratio of the magnetic field energy available to the rest mass energy of the particle. In the cold plasma limit, the Alfvén velocity, which is given by $V_A/c = \sqrt{\sigma/(\sigma + 1)}$, approaches the speed of light for the regime with a strong magnetic field of $\sigma \gg 1$. As the reconnection jet speed is known to be approximately equal to the Alfvén velocity regardless of the non-relativistic or relativistic nature of the reconnection, the reconnection jet speed almost reaches the speed of light when $\sigma \gg 1$. Therefore, from the flux frozen-in condition of $\mathbf{E} \sim -\mathbf{V}_A \times \mathbf{B}/c$, the magnitude of the reconnection electric field inside the reconnection jets may be almost the same as that of the reconnecting magnetic field, i.e., $E \sim B$. Then, efficient particle acceleration could operate in the reconnecting jets, for example, [64, 66, 67]. It is to be noted that if $E > B$, we can find a frame where $B = 0$ using Lorentz transformation.

To demonstrate the production of nonthermal particles in reconnection, Zenitani and Hoshino [68] and Jaroschek et al. [69] investigated the time evolution of a relativistic Harris current sheet with relativistic hot plasma temperatures of $T/mc^2 \sim O(0.1) - O(1)$ using 2D PIC simulations. In the Harris current sheet, the magnetic field \mathbf{B} in the x and y coordinate system is given by $B_x(y) = B_0 \tanh(y/\lambda)$, where λ is the thickness of the current sheet. The uniform plasma temperature $T(y) = \text{const.}$ and the plasma density $N(y) = N_0 \operatorname{sech}^2(y/\lambda) + N_b$ are set to maintain the pressure balance. The small background uniform density $N_b < N_0$ is often assumed to demonstrate continuous plasma injection from the lobe to the reconnection exhaust.

The relativistic simulations revealed the formation of a power-law energy spectrum, $N(\varepsilon) \propto \varepsilon^{-s}$ with $s \sim 1-2$, and it is argued that high-energy particles are preferentially generated in and around the X-type region, i.e., the magnetic diffusion region. The more the particles are accelerated, the longer they can stay in the diffusion region owing to the relativistic inertia effect. It has been discussed that the balance of the acceleration efficiency $d\varepsilon/dt = eEc$ and the loss rate of the accelerated particles from the diffusion region in the gyro-period $(dN(\varepsilon)/dt)/N(\varepsilon) \sim eBc/\varepsilon$ produces a power-law energy spectrum whose spectral index is $s \sim B/E$, where $N(\varepsilon)$ and ε represent the particle number and particle energy, respectively [68]. It should be noted that the gyro-period of the loss rate is inversely related to the particle's energy and its effect appears in relativistic plasmas.

After earlier studies using 2D PIC simulations, the 3D effect of reconnection has been investigated by focusing on both nonthermal particle acceleration and the formation of plasmoid/turbulence [70, 71]. In panels (a) and (b) of Fig. 11.3, we show the comparison between 2D and 3D simulations of the relativistic reconnection with $\sigma = 10^2$. The system sizes for the 2D and 3D simulations are 300×194 and $300 \times 300 \times 194$ per unit of ion inertia length (c/ω_{pi}), respectively. We can observe the stochastic generation and coalescence of the plasmoid in both 2D and 3D simulations, but in the 3D system, the drift-kink mode develops and interacts with the tearing/reconnection mode [72–75]. Then, the plasma sheet becomes much more dynamic and turbulent. In the bottom panel of the time evolution of the energy spectra, a hard energy spectrum with the index 1.35 is obtained in the 3D PIC

Fig. 11.3 Comparison of 2D and 3D relativistic reconnections with a high $\sigma = 10^2$. The top and middle panels of (a) and (b) are 2D and 3D current density structures, respectively. In the middle panel (b), the isosurface of the current density colored by $\mathbf{J} \cdot \mathbf{E}$ is depicted. Bottom panel (c) is the time evolution of the particle energy spectra obtained by 2D (dashed lines) and 3D (solid lines) simulations. The figure has been adopted from [70] with the permission of APS publishing



simulation, which essentially matches the 2D result. Sironi et al. [71] reported that a hard energy spectrum whose spectral index is greater than 2 can be obtained for a plasma sheet with a large $\sigma > 10$.

It is also interesting to mention that the particle acceleration can happen during the interaction of particles with the magnetic islands. For a long current sheet, many magnetic islands are formed in association with the tearing mode instability, and the islands can grow by the coalescence instability. The particles trapped around the magnetic island undergo first-order Fermi acceleration during the contraction within the magnetic island [76, 77].

In the relativistic regime, incoherent radiation losses such as synchrotron radiation and Compton radiation may play an important role in the dynamic evolution of the system. To include the effect of radiation loss in the PIC simulation, the Abraham-Lorentz equation with radiation drag can be implemented. In the Landau–Lifshitz form [78], the equation of motion is given by

$$\frac{d\mathbf{p}}{dt} = e \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right) + \mathbf{f}_{\text{rad}}, \quad (11.1)$$

with the radiation drag force

$$\mathbf{f}_{\text{rad}} = \frac{2}{3} \frac{e r_e}{c} \Gamma (d_t \mathbf{E} + \boldsymbol{\beta} \times d_t \mathbf{B}) + \frac{2}{3} r_e^2 (\mathbf{E} \times \mathbf{B} + \mathbf{B} \times (\mathbf{B} \times \boldsymbol{\beta}) + \mathbf{E}(\mathbf{E} \cdot \boldsymbol{\beta})) - \frac{2}{3} r_e^2 \Gamma^2 \boldsymbol{\beta} ((\mathbf{E} + \boldsymbol{\beta} \times \mathbf{B})^2 - (\mathbf{E} \cdot \boldsymbol{\beta})^2). \quad (11.2)$$

Here, the convective time derivative $d_t = \partial_t + \mathbf{v} \cdot \nabla$, three velocity $\boldsymbol{\beta} = \mathbf{v}/c$, Lorentz factor $\Gamma = 1/\sqrt{1 - \beta^2}$, and the classical electron radius $r_e = e^2/mc^2$. For a relativistic plasma with $\Gamma \gg 1$, the last term including Γ^2 obviously dominates, and we may neglect the other two terms. As long as the radiation drag force is smaller than the conventional electric and Lorentz forces, it can be numerically calculated as a post-process after the calculation of the electric and Lorentz forces.

Note that the radiation solved in the traditional PIC simulation is limited to the grid and time resolutions, while the incoherent radiation in association with the motion of an individual particle described by the drag force can cover from the low- to high-frequency ranges beyond the grid and time resolutions. A double counting of the radiation in the low-frequency range may be an open issue.

Jaroschek and Hoshino [79] studied the nonlinear time evolution of a radiation-dominated relativistic current sheet and reported that the current sheet collapses quickly owing to the reduction of gas pressure by the radiation loss, and a fast magnetic reconnection can happen. This effect gains importance for a large σ and/or a thin current sheet with a relativistic drift velocity. It is also interesting to calculate the radiation spectrum based on PIC simulation results, and it provides useful information for modeling astrophysical observations [80].

In addition to the fully relativistic reconnection reviewed above, magnetic reconnection in the trans-relativistic regime with non-relativistic ions and relativistic electrons has received great attention in high-energy astrophysics, as seen in accretion disk coronae in black hole systems [81, 82], in blazar jets [83], and so on. In these astrophysical phenomena, hadronic ion-electron plasmas are believed to dominate instead of positron-electron pair plasmas, and the energy partition between ions and electrons is central to understanding emission models that can be compared with observations. Rowan et al. [84] and Werner et al. [85] studied plasma heating and particle acceleration on the trans-relativistic reconnection for anti-parallel magnetic field geometry and demonstrated that energy deposition into ions is more efficient than electrons. Interestingly, a high ion-to-electron temperature is observed in the Earth's magnetosphere [86], and the preferential ion heating mechanism during reconnection has been investigated by PIC simulations as well [87, 88].

11.3.2 Magnetic Reconnection in Pulsar Winds

So far, we have reviewed the reconnection in an isolated system, but with the advances in computing power, the reconnection embedded in a global pulsar wind has been investigated recently using PIC simulations [89–92]. The pulsar

magnetosphere inside the light cylinder $R_{LC} = c/\Omega$, where Ω is the pulsar angular velocity, is theoretically believed to be occupied by strongly magnetized plasma owing to the strong magnetic field of the neutron star, whose surface magnetic field is approximately 10^{12} Gauss. From the observed synchrotron radiation in the Crab Nebula, it is argued that the pulsar wind nebulae are occupied by weakly magnetized plasma with nonthermal particles. As the magnetization parameter σ switches from $\sigma \gg 1$ in the pulsar magnetosphere to $\sigma \ll 1$ in the nebula, some energy conversion from the magnetic field energy to the particle kinetic energy has to occur during the wind expansion. This is the so-called σ problem, and the reconnection may provide an important clue to solve it.

When the magnetic dipole moment of the neutron star is not parallel to the rotation axis of the pulsar, the polarity of the magnetic field switches periodically every rotation, and the current sheet appears in a finite equatorial region, in a manner similar to the heliospheric current sheet.

Then, magnetic reconnection has received great attention as a plausible process of magnetic energy dissipation.

An important issue associated with reconnection in the pulsar wind is the expansion effect of the current sheet. With outward expansion, the magnitude of the magnetic field decays as $B \propto 1/r$, whereas the plasma density confined in the current sheet decreases as $N \propto 1/r^2$. From the Ampère's law, $J \propto Nv_d \propto B/L$, where v_d and L represent the drift velocity carrying the electric current and the thickness of the current sheet, respectively. The thickness of the current sheet L may be estimated to be nearly equal to or less than the size of the light cylinder $R_{LC} = c/\Omega$. If the thickness does not change during wind expansion, the drift velocity v_d increases with the distance from the pulsar. In addition to the effect of the relativistic hot plasma, the effect of the relativistic drift current should also be taken into account [93]. Moreover, the charged particles might not support the electric current maintained by the surrounding magnetic field, leading to the so-called charge starvation beyond some distance. The wind structure and the dissipation of magnetic energy during the course of wind expansion have been an interesting problem for more than half a century.

To solve the σ problem during the expansion of a pulsar wind, Cerutti et al. [89, 92] and Philippov et al. [90] investigated the interplay of the reconnection and pulsar wind expansion using PIC simulations. In Fig. 11.4, we show a two-dimensional slice of the three-dimensional pulsar global magnetosphere in the equatorial $r - \phi$ plane [92]. The stage of the well-developed pulsar wind structure is shown from the central pulsar up to a distance of 50 light cylinders. Initially, the magnetic field lines are assumed to be purely radial out of a split monopole, and they start to co-rotate with the central neutron star. Shortly after the start of the simulation, many current sheets under the Parker Archimedean spiral structure can be formed. The onset of magnetic reconnection and the formation of magnetic islands can be observed near the light cylinder, where the magnetic field topology changes from a closed magnetosphere inside the light cylinder into an open magnetosphere extended beyond the light cylinder. The behavior of the quick onset of reconnection is consistent with two-dimensional simulations [89]. One can observe

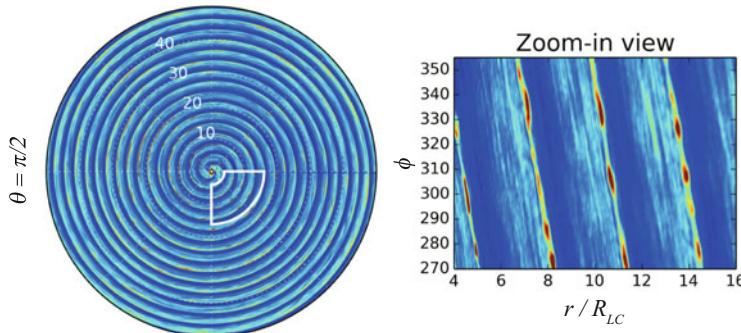


Fig. 11.4 2D slice of the 3D pulsar global magnetosphere in the equatorial $r - \phi$ plane. The plasma density $r^2 n$ is shown (red (blue) represents high (low) density regions); the distance from the central pulsar is normalized by the light cylinder, c/Ω , where Ω is the angular velocity of the pulsar. The right-hand panel is the enlarged view of the region indicated by the white fan-shaped box in the left-hand panel. The figure has been adopted from [92] as permitted by Astronomy and Astrophysics

that the tearing current sheets continue to develop with increasing distance from the pulsar, and the interaction among many tearing current sheets/magnetic islands enhances the magnetic energy dissipation. During this evolution, the broadening of the current sheet thickness was found to be proportional to r . In addition, in association with the magnetic energy dissipation, the production of nonthermal particles with a hard energy spectrum, which is consistent with high σ reconnection simulation studies in the isolated system [68–71], has been observed.

11.4 Magneto-Rotational Instability (MRI) in Collisionless Accretion Disks

Astroplasma phenomena around black holes, neutron stars, etc., are another important topic, and the main free energy available in these objects can often be attributed to the gravity and/or the rotation of gas. As one of the applications of astroplasma phenomena where gravity and rotation play an intriguing role in plasma dynamics, we now briefly review the magneto-rotation instability (MRI) in an accretion disk, which is regarded as an important process for understanding the outward transport of the angular momentum when mass accretion occurs inward [94]. Intense electromagnetic radiation is emitted when the gravitational energy is released.

For gas orbiting at a radius r around a central object under a Keplerian disk, the centrifugal force, $f_c \propto r\Omega^2$, and the gravitational force, $f_g \propto GM/r^2$, are balanced. The quantities Ω , M , and G are the angular velocity, mass of the central object, and gravitational constant, respectively. If the angular momentum, $L = r^2\Omega$, is conserved, the centrifugal force is proportional to $f_c \propto L^2/r^3$. Then, we find that the orbiting gas is stable against any perturbation. Therefore, a mechanism

that breaks the conservation of angular momentum is required for gas accretion. In the case of MRI, the magnetic tension force of the Alfvén wave modified in the gravitational-rotating system is involved in angular momentum transport [95, 96].

The angular momentum transport has been extensively studied in the framework of the MHD approximation [95–99], and a weakly magnetized accretion disk with outwardly decreasing angular velocity is known to be unstable in MRI. During the temporal evolution, the weak initial magnetic field can be amplified by the action of the MRI dynamo owing to the differential motion of the accretion disk. The amplified magnetic field is subject to magnetic reconnection, and the saturation of the MRI dynamo can be determined by the balance between the growth of the magnetic field by the MRI dynamo and the magnetic dissipation by reconnection. Based on many MHD simulation results, it is argued that MRI can develop into the plasma state with the thermal energy and magnetic field energy almost equipartitioned, namely, the accretion disk with $\beta \sim O(1)$ can be realized. MRI generates enough MHD turbulence and Maxwell stress to facilitate angular momentum transport and mass accretion.

So far, the physics of MRI have been successfully understood in the MHD framework, whereas the accretion disk around a black hole is believed to be in a collisionless plasma state. In fact, from the supermassive black hole at Sagittarius A*, the non-equilibrium temperature between protons and electrons as well as nonthermal high-energy particles are observed, for example, [100–102]. Therefore, it is important to understand the nonlinear behavior of MRI and the production of nonthermal particles in collisionless systems. It is also interesting to self-consistently investigate the saturation level of the MRI dynamo in a collisionless system, which may be sensitive to the kinetic energy dissipation process by magnetic reconnection and turbulence.

To understand the kinetic dynamics of MRI, its PIC simulations have been performed in a relatively small local system rotating with angular velocity Ω_0 at a distance r_0 from the center of gravity. As the MRI is classified as a local instability, PIC simulation in a local box would be sufficient for capturing the fundamental process involved.

A few assumptions and implementations of the simulation algorithm were adopted for studying the local system. First, we need to add the Coriolis, centrifugal, and gravitational forces to the equations of motion. We used the tidal expansion of the effective potential with a constant $q = -\partial \ln \Omega / \partial \ln r$ at radius r_0 , where q is 3/2 for a Keplerian disk, by assuming that the force balance of $GM/r^2 = r\Omega(r)^2$ is satisfied, that is, the tidal expansion in a local system rotating with angular velocity Ω_0 can be given by

$$r\Omega_0^2 - \frac{GM}{r^2} = r\Omega_0^2 - r\Omega(r)^2 \simeq r\Omega_0^2 - r \left(\Omega_0 + \frac{\partial \Omega_0}{\partial r}x \right)^2 \simeq 2q\Omega_0^2x,$$

where $x = r - r_0$. Then, the equation of motion is given by

$$\frac{d\mathbf{p}}{dt} = e \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right) - m\Gamma(2\Omega_0 \times \mathbf{v} - 2q\Omega_0^2 x \mathbf{e}_x), \quad (11.3)$$

where $\Gamma = \sqrt{1 + (p/mc)^2}$. We have neglected the vertical component of gravity.

Instead of the local cylindrical coordinate system, we introduce the Cartesian coordinates x , y , and z , which correspond to the radial direction $x = r$, azimuthal direction $y = r_0\phi$, and direction parallel to the rotation axis z , respectively. We assume that the size of the simulation box is significantly smaller than its distance from the center of the disk.

The boundary condition of x , y , and z is basically the periodic boundary condition, and we employed the shearing box approximation owing to the differential rotation in the azimuthal direction [103]. Physical quantities are obtained by the Lorentz transformation based on the velocity differences at the boundary, and the differential rotation velocity $r\Omega_0(r_0)$ is calibrated at the boundary.

Another important technical issue in PIC simulations is that the local rotating frame with the angular velocity Ω_0 is not an inertial frame, and hence, some additional correctional terms should be included in the set of Maxwell's equations [104]. If the angular velocity is much less than the speed of light, the correctional terms are negligible. This is discussed in detail in [105, 106].

The time evolution of the MRIs, obtained by 3D PIC simulations, is shown in Fig. 11.5. For simplicity, a pair plasma was assumed. For the initial condition, a drifting Maxwellian velocity distribution function was assumed in the local rotating frame with the angular velocity $\Omega_0(r_0)$, and the drift velocity in the y direction $v_y(x)$ was given by $v_y(x) = r(\Omega(r) - \Omega_0(r_0)) \simeq -q\Omega_0(r_0)x$. Other setup parameters are discussed in [107]. Panel (a) shows the magnetic field lines (green lines) in the initial state, which are parallel to the rotation axis z . The angular velocities of the Keplerian differential motion can be seen as the color contour in the slice cut at $Y = 1.9$. The special scale is normalized by the most unstable wavelength of MRI, $2\pi V_A/\Omega_0$. The reddish (bluish) region corresponds to a positive (negative) toroidal velocity. In the temporal evolution shown in Panel (b), at the time stage $T = 6.89$ normalized by the orbital rotation period $\Omega_0/2\pi$, the magnetic field lines parallel to the rotation axis start to be distorted in the radial x direction, and they stretch out toward the toroidal y direction because of the frozen-in motion between the magnetic field and the Keplerian differential motion. Panel (c) has the same time stage as Panel (b), but the high-density regions are shown by the red curved planes, where inward- and outward-flowing streams in association with a high plasma density and strong electric current are formed. This is the so-called channel flow. Panel (d) ($T = 7.18$) corresponds to the early nonlinear stage, whereas Panel (e) is the late nonlinear stage. The amplified magnetic field of the MRI dynamo/Keplerian differential motion is balanced by the dissipation of the collisionless magnetic reconnection. In PIC simulations, the collisionless resistivity caused by the particle inertia can be self-consistently evaluated in the magnetic reconnection process. The dynamic breakage of the channel flow/current sheet can

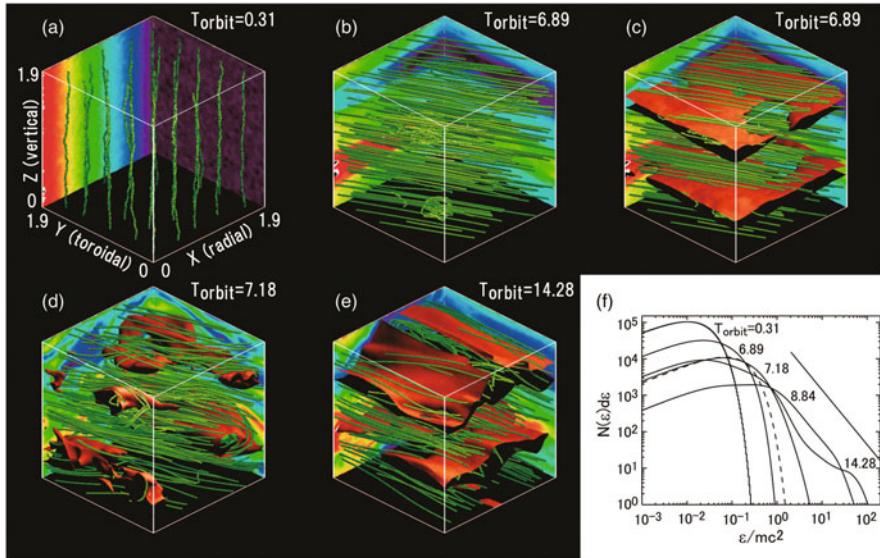


Fig. 11.5 Time evolution of the magnetorotational instability in the PIC simulation. The magnetic field lines (green lines) and angular velocities at $X = 1.9$ and $Y = 1.9$ (color contour) are shown in panels (a)–(e). The high-density regions are also shown as reddish curved planes in panels (c)–(e). Panel (f) shows the time evolution of the energy spectra with the formation of nonthermal particles. The figure has been adopted from [107] with the permission of APS publishing

be observed, and the repeated reformation and destruction of the channel flow occur intermittently owing to the competition between the MRI dynamo action and the magnetic reconnection. These nonlinear time evolutions seen in the PIC simulation are basically the same as those demonstrated in the MHD simulations.

However, an intriguing feature of the collisionless accretion disks is the production of nonthermal particles. In Panel (f), which shows the time evolution of the energy spectra, the horizontal axis is the particle energy ε normalized by the rest mass energy mc^2 , and the vertical axis is the number density $N(\varepsilon)$. In the growth phase of the MRI before $T = 6.79$, the plasma is gradually heated from the initial cold Maxwellian plasma, but after the onset of reconnection at $T = 7.18$, nonthermal high-energy tails can be clearly observed. The dashed line, for reference, is a thermal Maxwellian spectrum with $T/mc^2 = 0.121$. As time passes, the nonthermal tail continues to grow, and the spectrum can be approximated by a single power-law function with $N(\varepsilon) \propto \varepsilon^{-3/2}$. The formation of nonthermal particles is attributed to magnetic reconnection inside the channel flow [105–107].

Another important issue is understanding whether the angular momentum transport at the kinetic level is enhanced compared to that discussed in the MHD framework. If the dissipation of the magnetic field is suppressed in the collisionless system, the angular momentum transport is expected to be enhanced. This is because the saturation level of the magnetic field is determined by the balance between the

dynamic amplification and dissipation of the magnetic field under the MRI dynamo process and the magnetic reconnection, as well as the increase in angular momentum transport with an increase in the magnitude of the magnetic field in the accretion disk.

Based on the kinetic MRI simulation, it has been discussed that the pressure anisotropy generated during the collisionless MRI can provide a much more efficient angular momentum transport with one order of magnitude larger than that discussed in the MHD framework [107]. In the equation of state, using the double adiabatic approximation [108], it is known that $p_{\parallel} > p_{\perp}$ is produced in the channel flow during magnetic reconnection [107]. Then the magnetic tension force of MHD waves can be modified, and the rate of the angular momentum transport is expected to change. It is claimed that the anisotropic stress tensor $(p_{\parallel} - p_{\perp})B_x B_y / B^2$ has almost the same contribution as the Maxwell stress tensor $B_x B_y / 4\pi$ in the kinetic models [107, 109, 110]. Furthermore, the enhancement of the parallel pressure $p_{\parallel} > p_{\perp}$ by reconnection suppresses the onset of subsequent reconnection. This, in turn, causes high magnetic-field saturation and enhancement of the stress tensors in collisionless plasma [107]. It is noteworthy that a similar conclusion has been obtained in the kinetic MHD simulation that can handle six elements of the pressure tensor [110, 111]. It would be necessary to study the angular momentum transport in a fully turbulent state with a large-scale PIC simulation.

11.5 Summary

The advances in supercomputer capacities and numerical algorithms have led to the recognition of full particle-in-cell simulations of astroplasma phenomena as useful tools for investigating not only the kinetic plasma instabilities and transports in a local system but also the global phenomena involving the coupling of macroscale and microscale kinetics. In fact, in addition to the several simulations introduced in this chapter, many intriguing PIC simulations, such as the black hole magnetosphere and its jet launching, have been studied, e.g., [112]. New physical processes and effects such as the radiation effect [79], pair production [113], gravitationally rotating effect [107], and general relativistic effect [90] are now implemented in conventional PIC simulations. However, 3D simulation studies with a realistic mass ratio of ions to electrons are still challenging compared with MHD and hybrid simulations. Furthermore, to study a wide dynamic energy range in high-energy astrophysical phenomena, it is necessary to increase the system size as much as possible. Accordingly, it is also necessary to calculate a longer evolution time. The rapid progress in supercomputers will overcome these difficulties in the near future, leading to a better understanding of the plasma universe.

Acknowledgments This work has been funded by JSPS Grant-in-Aid for Challenging Exploratory Research Grant Nos. 18K18748 and 20K20908 and was also supported in part by the National Science Foundation under Grant No. NSF-PHY-1748958. MH thanks the program of High Energy Plasma Phenomena in Astrophysics 2019 at the Kavli Institute for Theoretical Physics.

References

1. K. Koyama, R. Petre, E.V. Gotthelf, et al., Evidence for shock acceleration of high-energy electrons in the supernova remnant SN1006. *Nature* **378**, 255 (1995)
2. P.F. Winkler, K.S. Long, *Astrophys. J.* **491**, 829 (1997)
3. Y. Lyubarsky, J.G. Kirk, *Astrophys. J.* **547**, 437 (2001)
4. J.G. Kirk, O. Skjæraasen, *Astrophys. J.* **591**, 366 (2003)
5. M.J. Rees, P. Meszaros, *Mon. Not. R. Astron. Soc.* **258**, 41 (1992)
6. H.C. Spruit, F. Daigne, G. Drenkhahn, *Astron. Astrophys.* **369**, 694 (2001)
7. D. Giannios, D.A. Uzdensky, M.C. Begelman, *Mon. Not. R. Astron. Soc.* **395**, L29 (2009)
8. D.R. Lorimer, M. Bailes, M.A. McLaughlin, et al., *Science* **318**, 777 (2007)
9. M. Pohl, M. Hoshino, J. Niemiec, *Progr. Part. Nuclear Phys.* **111**, 103751 (2020). <https://doi.org/10.1016/j.ppnp.2019.103751>
10. A.R. Bell, *Mon. Not. R. Astron. Soc.* **182**, 147 (1978)
11. R.D. Blandford, J.P. Ostriker, *Astrophys. J.* **221**, L29 (1978)
12. A. Levinson, *Mon. Not. R. Astron. Soc.* **278**, 1018 (1996)
13. T. Amano, M. Hoshino, *Astrophys. J.* **661**, 190 (2007)
14. R.Z. Sagdeev, *Rev. Plasma Phys.* **4**, 23 (1966)
15. C.S. Wu, D. Winske, Y.M. Zhou, et al., *Space Sci. Rev.* **37**, 63 (1984)
16. M.M. Leroy, D. Winske, C.C. Goodrich, et al., *J. Geophys. Res.* **87**, 5081 (1982)
17. H. Naitou, T. Kamimura, S. Tokuda, *J. Comput. Phys.* **33**, 86 (1979). [https://doi.org/10.1016/0021-9991\(79\)90029-9](https://doi.org/10.1016/0021-9991(79)90029-9)
18. M.E. Dieckmann, S.C. Chapman, K.G. McClements, et al., *Astron. Astrophys.* **356**, 377 (2000)
19. N. Shimada, M. Hoshino, *Astrophys. J.* **543**, L67 (2000)
20. K.G. McClements, M.E. Dieckmann, A. Ynnerman, et al., *Phys. Rev. Lett.* **87**, 255002 (2001)
21. B. Lembege, J. Giacalone, M. Scholer, et al., *Space Sci. Rev.* **110**, 161 (2004)
22. A. Marcowith, A. Bret, A. Bykov, et al., *Rep. Progr. Phys.* **79**, 046901 (2016)
23. A. Bohdan, J. Niemiec, M. Pohl, et al., *Astrophys. J.* **878**, 5 (2019)
24. Y. Matsumoto, T. Amano, T.N. Kato, et al., *Phys. Rev. Lett.* **119**, 105101 (2017)
25. O. Buneman, *Phys. Rev. Lett.* **1**, 8 (1958)
26. E.S. Weibel, *Phys. Rev. Lett.* **2**, 83 (1959)
27. Y. Matsumoto, T. Amano, M. Hoshino, *Astrophys. J.* **755**, 109 (2012)
28. M. Hoshino, N. Shimada, *Astrophys. J.* **572**, 880 (2002)
29. A. Bohdan, J. Niemiec, M. Pohl, et al., *Astrophys. J.* **885**, 10 (2019). <https://doi.org/10.3847/1538-4357/ab43cf>
30. Y. Matsumoto, T. Amano, T.N. Kato, et al., *Science* **347**, 974 (2015)
31. R.Z. Sagdeev, V.D. Shapiro, *Soviet J. Exper. Theor. Phys. Lett.* **17**, 279 (1973)
32. M. Hoshino, *Progr. Theor. Phys. Suppl.* **143**, 149 (2001)
33. A. Bohdan, M. Pohl, J. Niemiec, et al., *Astrophys. J.* **893**, 6 (2020). <https://doi.org/10.3847/1538-4357/ab7cd6>
34. M. Hoshino, *Phys. Rev. Lett.* **108**, 135003 (2012)
35. M. Hoshino, *Science* **299**, 834 (2003)
36. A. Bohdan, M. Pohl, J. Niemiec, et al., *Astrophys. J.* **904**, 12 (2020). <https://doi.org/10.3847/1538-4357/abbc19>
37. A.B. Langdon, J. Arons, C.E. Max, *Phys. Rev. Lett.* **61**, 779 (1988)
38. M. Hoshino, J. Arons, Y.A. Gallant, et al., *Astrophys. J.* **390**, 454 (1992)

39. Y.A. Gallant, M. Hoshino, A.B. Langdon, et al., *Astrophys. J.* **391**, 73 (1992)
40. Y. Lyubarsky, *Astrophys. J.* **652**, 1297 (2006)
41. M. Hoshino, *Astrophys. J.* **672**, 940 (2008)
42. A. Spitkovsky, *Astrophys. J.* **682**, L5 (2008)
43. T.N. Kato, H. Takabe, *Phys. Plasmas* **17**, 032114 (2010)
44. L. Sironi, A. Spitkovsky, *Astrophys. J.* **726**, 75 (2011)
45. M. Hoshino, J. Arons, *Phys. Fluids B* **3**, 818 (1991)
46. W.L. Kruer, *Proc. SPIE* **913**, 2 (1988). <https://doi.org/10.1117/12.965116>
47. T. Tajima, J.M. Dawson, *Phys. Rev. Lett.* **43**, 267 (1979)
48. K. Mima, K. Nishikawa, *Basic Plasma Physics: Selected Chapters, Handbook of Plasma Physics*, vol. 1 (North-Holland, Amsterdam, 1984), p. 451
49. Y. Kuramitsu, N. Nakanii, K. Kondo, et al., *Phys. Plasmas* **18**, 010701 (2011)
50. M. Iwamoto, T. Amano, M. Hoshino, et al., *Astrophys. J.* **840**, 52 (2017)
51. M. Iwamoto, T. Amano, M. Hoshino, et al., *Astrophys. J.* **858**, 93 (2018)
52. B.B. Godfrey, *J. Comput. Phys.* **15**, 504 (1974)
53. S.F. Martins, R.A. Fonseca, L.O. Silva, et al., *Comput. Phys. Commun.* **181**, 869 (2010)
54. J.-L. Vay, C.G.R. Geddes, C. Benedetti, et al., *American Institute of Physics Conference Series* (2010), p. 244
55. J.-L. Vay, C.G.R. Geddes, E. Cormier-Michel, et al., *J. Comput. Phys.* **230**, 5908 (2011)
56. B.B. Godfrey, J.-L. Vay, *J. Comput. Phys.* **248**, 33 (2013)
57. D.-Y. Na, J.L. Nicolini, R. Lee, et al., *J. Comput. Phys.* **402**, 108880 (2020). <https://doi.org/10.1016/j.jcp.2019.108880>
58. N. Ikeya, Y. Matsumoto, *Publ. Astron. Soc. Jpn.* **67**, 64 (2015)
59. M. Iwamoto, T. Amano, M. Hoshino, et al., *Astrophys. J.* **883**, L35 (2019). <https://doi.org/10.3847/2041-8213/ab4265>
60. M.V. Medvedev, A. Loeb, *Astrophys. J.* **526**, 697 (1999)
61. B.D. Metzger, B. Margalit, L. Sironi, *Mon. Not. R. Astron. Soc.* **485**, 4091 (2019)
62. J. Giacalone, *Astrophys. J.* **609**, 452 (2004). <https://doi.org/10.1086/421043>
63. R. Blandford, Y. Yuan, M. Hoshino, et al., *Space Sci. Rev.* **207**, 291 (2017)
64. M. Hoshino, Y. Lyubarsky, *Space Sci. Rev.* **173**, 521 (2012)
65. F. Guo, Liu, Y.-H., X. Li, et al., *Phys. Plasmas* **27**, 080501 (2020). <https://doi.org/10.1063/5.0012094>
66. M.M. Romanova, R.V.E. Lovelace, *Astron. Astrophys.* **262**, 26 (1992)
67. D.A. Larrabee, R.V.E. Lovelace, M.M. Romanova, *Astrophys. J.* **586**, 72 (2003)
68. S. Zenitani, M. Hoshino, *Astrophys. J.* **562**, L63 (2001)
69. C.H. Jaroschek, H. Lesch, R.A. Treumann, *Astrophys. J.* **605**, L9 (2004)
70. F. Guo, H. Li, W. Daughton, et al., *Phys. Rev. Lett.* **113**, 155005 (2014)
71. L. Sironi, A. Spitkovsky, *Astrophys. J.* **783**, L21 (2014)
72. J. Büchner, J.-P. Kuska, *Ann. Geophys.* **17**, 604 (1999)
73. W. Daughton, *Phys. Plasmas* **6**, 1329 (1999)
74. S. Zenitani, M. Hoshino, *Phys. Rev. Lett.* **95**, 095001 (2005)
75. S. Zenitani, M. Hoshino, *Astrophys. J.* **670**, 702 (2007)
76. J.F. Drake, M. Swisdak, H. Che, et al., *Nature* **443**, 553 (2006). <https://doi.org/10.1038/nature05116>
77. F. Guo, X. Li, W. Daughton, et al., *Astrophys. J.* **879**, L23 (2019). <https://doi.org/10.3847/2041-8213/ab2a15>
78. L.D. Landau, E.M. Lifshitz, *Course of Theoretical Physics—Pergamon International Library of Science* (Butterworth-Heinemann, Oxford, 1975)
79. C.H. Jaroschek, M. Hoshino, *Phys. Rev. Lett.* **103**, 075002 (2009)
80. D.A. Uzdensky, *Magnetic Reconnection: Concepts and Applications* (Springer, Cham, 2016), p. 473
81. R.A. Remillard, J.E. Mc Clintock, *Ann. Rev. Astron. Astrophys.* **44**, 49 (2006). <https://doi.org/10.1146/annurev.astro.44.051905.092532>

82. C. Done, M. Gierliński, A. Kubota, *Astron. Astrophys. Rev.* **15**, 1 (2007). <https://doi.org/10.1007/s00159-007-0006-1>
83. G. Madejski (Greg) , M. Sikora, *Ann. Rev. Astron. Astrophys.* **54**, 725 (2016). <https://doi.org/10.1146/annurev-astro-081913-040044>
84. M.E. Rowan, L. Sironi, R. Narayan, *Astrophys. J.* **850**, 29 (2017). <https://doi.org/10.3847/1538-4357/aa9380>
85. G.R. Werner, D.A. Uzdensky, M.C. Begelman, et al., *Mon. Not. R. Astron. Soc.* **473**, 4840 (2018). <https://doi.org/10.1093/mnras/stx2530>
86. W. Baumjohann, G. Paschmann, C.A. Cattell, *J. Geophys. Res.* **94**, 6597 (1989). <https://doi.org/10.1029/JA094iA06p06597>
87. C.C. Haggerty, M.A. Shay, J.F. Drake, et al., *Geophys. Res. Lett.* **42**, 9657 (2015). <https://doi.org/10.1002/2015GL065961>
88. M. Hoshino, *Astrophys. J.* **868**, L18 (2018). <https://doi.org/10.3847/2041-8213/aaef3a>
89. B. Cerutti, A. Philippov, K. Parfrey, et al., *Mon. Not. R. Astron. Soc.* **448**, 606 (2015)
90. A.A. Philippov, A. Spitkovsky, B. Cerutti, *Astrophys. J.* **801**, L19 (2015)
91. C. Kalapotharakos, G. Brambilla, A. Timokhin, et al., *Astrophys. J.* **857**, 44 (2018). <https://doi.org/10.3847/1538-4357/aab550>
92. B. Cerutti, A.A. Philippov, G. Dubus, *Astron. Astrophys.* **642**, A204 (2020). <https://doi.org/10.1051/0004-6361/202038618>
93. M. Hoshino, *Astrophys. J.* **900**, 66 (2020). <https://doi.org/10.3847/1538-4357/aba59d>
94. D. Lynden-Bell, J.E. Pringle, *Mon. Not. R. Astron. Soc.* **168**, 603 (1974). <https://doi.org/10.1093/mnras/168.3.603>
95. S.A. Balbus, J.F. Hawley, *Astrophys. J.* **376**, 214 (1991)
96. S.A. Balbus, J.F. Hawley, *Astrophys. J.* **400**, 610 (1992)
97. R. Matsumoto, T. Tajima, *Astrophys. J.* **445**, 767 (1995)
98. J.M. Stone, J.F. Hawley, C.F. Gammie, et al., *Astrophys. J.* **463**, 656 (1996)
99. T. Sano, S.-I. Inutsuka, N. J. Turner, et al., *Astrophys. J.* **605**, 321 (2004)
100. R. Narayan, I. Yi, R. Mahadevan, *Nature* **374**, 623 (1995)
101. H. Falcke, S. Markoff, *Astron. Astrophys.* **362**, 113 (2000)
102. S.S. Doeleman, J. Weintraub, A.E.E. Rogers, et al., *Nature* **455**, 78 (2008)
103. J.F. Hawley, C.F. Gammie, S.A. Balbus, *Astrophys. J.* **440**, 742 (1995)
104. L.I. Schiff, *Proc. Nat. Acad. Sci.* **25**, 391 (1939)
105. M.A. Riquelme, E. Quataert, P. Sharma, et al., *Astrophys. J.* **755**, 50 (2012)
106. M. Hoshino, *Astrophys. J.* **773**, 118 (2013)
107. M. Hoshino, *Phys. Rev. Lett.* **114**, 061101 (2015)
108. G.F. Chew, M.L. Goldberger, F.E. Low, *Proc. R. Soc. Lond. Ser. A* **236**, 112 (1956)
109. P. Sharma, G.W. Hammett, E. Quataert, et al., *Astrophys. J.* **637**, 952 (2006)
110. K. Hirabayashi, M. Hoshino, *Astrophys. J.* **842**, 36 (2017)
111. K. Hirabayashi, M. Hoshino, T. Amano, *J. Comput. Phys.* **327**, 851 (2016)
112. K. Parfrey, A. Philippov, B. Cerutti, *Phys. Rev. Lett.* **122**, 035101 (2019)
113. A.Y. Chen, A.M. Beloborodov, *Astrophys. J.* **795**, L22 (2014)
114. Y. Matsumoto, T. Amano, M. Hoshino, *Phys. Rev. Lett.* **111**, 215003 (2013)

Part III

Introduction to Advanced New Algorithms and Developments for Future Simulations

Although methods of space and astrophysical plasma simulation have been utilized since first powerful computers became available in the 1960s, optimum algorithms and codes are continuously under development. To illustrate this, the third part of this book provides examples of the development of new, advanced numerical techniques as for the use of higher-order schemes to investigate turbulent processes (Chap. 12) and for the development of new, artificial-intelligence, agent-based technologies for the better understanding of essentially multi-scale plasma and other multi-dimensional systems (Chap. 13).



Higher-Order Magnetohydrodynamic Simulations

12

Jean-Mathieu Teissier and Wolf-Christian Müller

Abstract

In this chapter, we aim to present the basic techniques necessary to go beyond the widely accepted paradigm of second-order numerics. We specifically focus on finite-volume schemes for hyperbolic conservation laws occurring in fluid approximations such as the equations of ideal magnetohydrodynamics or the Euler equations of gas dynamics. For the sake of clarity, a simple fourth-order ideal magnetohydrodynamic (MHD) solver which allows to simulate strongly shocked systems serves as an instructive example. Issues that only or mainly arise in the world of higher-order numerics are given specific focus. Alternative algorithms as well as refinements and improvements are discussed and are referenced to in the literature. As an example of application, some results on decaying compressible turbulence are presented.

12.1 Introduction

The nonlinear, typically multiscale and multi-physics character of, e.g., the turbulent dynamics of the solar wind, the solar corona, or planetary magnetospheres, requires reliable efficient and accurate numerical simulations as integral part of theoretical research in basically all domains of space-plasma physics. In this context, the description of plasmas as single- or multi-fluid systems has proven to be a powerful approximation (cf. Chap. 1). This is particularly true in settings where the scales of interest are beyond the realm of kinetic plasma modeling. A reflection of the level of physical simplification entailed in the fluid approximation is the

J.-M. Teissier · W.-C. Müller (✉)

Center for Astronomy and Astrophysics, Technical University Berlin, Berlin, Germany
e-mail: jm.teissier@astro.physik.tu-berlin.de; wolf-christian.mueller@tu-berlin.de

reduction of the phase-space particle distribution function to its lowest-order velocity moments: mass, momentum, and energy. Neglecting non-ideal dissipative processes, those quantities are strictly conserved. This assumed ideality of plasma dynamics is a common approximation given the low levels of collisionality of many space plasmas. As a consequence, astrophysical numerics typically has to deal with nonlinear conservation laws of the abovementioned quantities. Those partial differential equations are of hyperbolic type which means that the character of their solution is mainly determined by the speeds of signal propagation associated with the underlying physical processes such as fluid advection or wave propagation. If the fluid velocity is of the order of the local sound speed or exceeds it—a rather common situation in space—the solution can lose its regularity and propagating discontinuities, shock fronts, emerge.

The numerical challenge consists of dealing with the local loss of differentiability while still guaranteeing the conservation properties of the fluid equations and thus ensuring the physically correct evolution of the system. Different strategies of spatial discretization are available to this end, of which we mention only a few classical examples (for a somewhat orthogonal approach to the proper representation of the solution, see Chap. 7 on adaptive mesh refinement): the finite-difference approach representing the physical fields as a regular grid of point values, the finite-volume way based on volume averages of the fluid observables over arbitrarily shaped grid cells, and the finite-element technique which generalizes the finite-volume ansatz to a more complex and flexible framework for the representation of dependent variables and differential equations on a grid cell.

In this contribution as well as in many numerical investigations during the last decades, the finite-volume technique is chosen as a reasonable compromise between flexibility (compared to finite differences) and complexity (compared to finite elements). It evolves the cell averages of the conserved quantities via their fluxes on the cell boundaries and, consequently, is conservative by construction.

We present in Sect. 12.2 the general numerical framework: the equations to be solved, the motivation behind higher-order numerics, and the chosen discretization. Sections 12.3, 12.4, and 12.5 are concerned with the practical implementation of the scheme: the computation of higher-order fluxes, the time integration, and how strong shocks can be handled, respectively. The numerical scheme is validated through several tests in Sect. 12.6. Concluding remarks are given in Sect. 12.7.

12.2 General Numerical Framework

12.2.1 Basic Equations

We are interested in solving the ideal magnetohydrodynamic (MHD) equations assuming adiabatic thermodynamics. Since we present here a finite-volume scheme that employs a constrained transport method to evolve the magnetic field \mathbf{B} , it is convenient to express the differential evolution laws as two systems of equations. The hydrodynamic variables, mass density ρ , velocity \mathbf{v} , and total energy density e

(the sum of internal, kinetic, and magnetic energy), are governed by the continuity equation, the momentum balance and the energy equation. They are written here in conservative form:

$$\partial_t \rho = -\nabla \cdot (\rho \mathbf{v}), \quad (12.1)$$

$$\partial_t (\rho \mathbf{v}) = -\nabla \cdot \left(\rho \mathbf{v} \mathbf{v}^T + \left(p + \frac{1}{2} |\mathbf{B}|^2 \right) \mathbf{I} - \mathbf{B} \mathbf{B}^T \right), \quad (12.2)$$

$$\partial_t e = -\nabla \cdot \left(\left(e + p + \frac{1}{2} |\mathbf{B}|^2 \right) \mathbf{v} - (\mathbf{v} \cdot \mathbf{B}) \mathbf{B} \right), \quad (12.3)$$

with p the thermal pressure and \mathbf{I} the 3×3 identity matrix so that $\nabla \cdot (p + \frac{1}{2} |\mathbf{B}|^2) \mathbf{I} = \nabla(p + \frac{1}{2} |\mathbf{B}|^2)$. Using the adiabatic equation of state for an ideal gas, the internal energy is $\frac{p}{\gamma-1}$ with γ the ratio of specific heats, which means that p can be deduced through:

$$p = (\gamma - 1) \left(e - \frac{1}{2} \rho |\mathbf{v}|^2 - \frac{1}{2} |\mathbf{B}|^2 \right). \quad (12.4)$$

In the isothermal case, $\gamma = 1$ and the pressure is determined by $p = \rho c_s^2$ with constant sound speed c_s (when assuming a constant initial temperature). In that case, no equation for e needs to be solved.

As for the magnetic field, its evolution is governed by:

$$\partial_t \mathbf{B} = -\nabla \times \mathbf{E}, \quad (12.5)$$

with $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$ the electric field. In addition to the above equations, the absence of magnetic monopoles requires the solenoidality of the magnetic field at all times:

$$\nabla \cdot \mathbf{B} = 0. \quad (12.6)$$

12.2.2 Motivation for Using Higher-Order Schemes

Numerical schemes whose discretization error is of second order in the grid scale are often combined with more precise Riemann solvers that are required in finite-volume schemes to calculate the fluxes over the faces of a grid cell (see Sect. 12.3.3). An alternative way to reduce numerical dissipation are higher-order schemes which increase the order of the overall discretization error to accelerate the convergence toward the reference solution in continuous space and time as the numerical resolution is increased. In this chapter, we qualify a scheme as “higher-order” if its discretization order is strictly greater than 3 and of “lower-order” otherwise. Higher-order schemes by definition represent the solution by considering more terms of the associated Taylor series at a specific location on the numerical grid. Consequently,

they achieve a more accurate and therefore less dissipative approximation of the reference solution. Although they generally have a higher computational cost at a given resolution as compared to a lower-order scheme, a higher-order algorithm allows to obtain the same solution quality at lower numerical resolution, provided the method remains stable. Therefore, and in particular in more than one spatial dimension, their employment allows a significant reduction of computational cost. Section 12.6 corroborates this statement by comparison of fourth-order and lower-order schemes.

12.2.3 Finite-Volume

In a finite-volume approach, spatial cell averages are evolved in time. We consider hence the time evolution of the cell average:

$$\mathbf{U}_{i,j,k} = \frac{1}{\Delta x \Delta y \Delta z} \int_{\Omega_{i,j,k}} \mathbf{u}(x, y, z) dx dy dz, \quad (12.7)$$

with $\mathbf{u} = (\rho, \rho v_x, \rho v_y, \rho v_z, e)$ a vector containing the hydrodynamical quantities and \mathbf{U} the corresponding cell averages over the cell $\Omega_{i,j,k} = [x_i - \frac{\Delta x}{2}, x_i + \frac{\Delta x}{2}] \times [y_j - \frac{\Delta y}{2}, y_j + \frac{\Delta y}{2}] \times [z_k - \frac{\Delta z}{2}, z_k + \frac{\Delta z}{2}]$. We assume for the sake of simplicity a Cartesian coordinate system where $(x_i = (i+1/2)\Delta x, y_j = (j+1/2)\Delta y, z_k = (k+1/2)\Delta z)$ is the center of the cell indexed by (i, j, k) and $\Delta x, \Delta y, \Delta z$ are constant, not necessarily equal grid sizes. Using Eqs.(12.1)–(12.3) and after application of Gauss' theorem, we obtain:

$$\partial_t \mathbf{U}_{i,j,k} = -\frac{D_x(\mathbf{F}_{i,j,k}^{A,x})}{\Delta x} - \frac{D_y(\mathbf{F}_{i,j,k}^{A,y})}{\Delta y} - \frac{D_z(\mathbf{F}_{i,j,k}^{A,z})}{\Delta z}, \quad (12.8)$$

where we introduced in order to have concise notations a difference operator:

$$D_x(F_{i,j,k}) = F_{i+1/2,j,k} - F_{i-1/2,j,k} \quad (12.9)$$

for any quantity F , and similarly $D_y(F_{i,j,k}) = F_{i,j+1/2,k} - F_{i,j-1/2,k}$ and $D_z(F_{i,j,k}) = F_{i,j,k+1/2} - F_{i,j,k-1/2}$. The fluxes $\mathbf{F}^{A,x}, \mathbf{F}^{A,y}$, and $\mathbf{F}^{A,z}$ along the $\mathbf{x}-$, $\mathbf{y}-$, and $\mathbf{z}-$ directions, respectively, are area-averaged (see Fig. 12.1), that is:

$$\mathbf{F}_{i\pm 1/2,j,k}^{A,x} = \frac{1}{\Delta y \Delta z} \int_{A_{j,k}^x} \mathbf{f}^x(x_{i\pm 1/2}, y, z) dy dz, \quad (12.10)$$

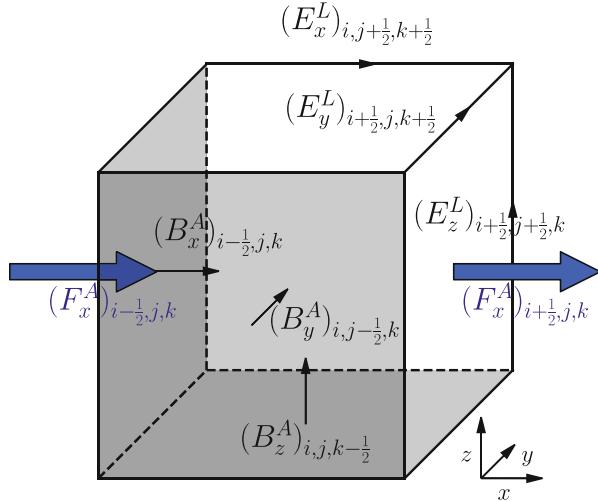


Fig. 12.1 Definition of the magnetic quantities and the fluxes in the cell $\Omega_{i,j,k}$. Only the flux in the x -direction is shown; the ones in the other directions are defined analogously

where $A_{j,k}^x = [y_j - \frac{\Delta y}{2}, y_j + \frac{\Delta y}{2}] \times [z_k - \frac{\Delta z}{2}, z_k + \frac{\Delta z}{2}]$ and similarly for $\mathbf{F}^{A,y}$ and $\mathbf{F}^{A,z}$. The fluxes \mathbf{f}^x , \mathbf{f}^y , and \mathbf{f}^z can be derived from Eqs. (12.1)–(12.3), for example:

$$\mathbf{f}^x = \begin{pmatrix} \rho v_x \\ \rho v_x^2 + p + |\mathbf{B}|^2/2 - B_x^2 \\ \rho v_x v_y - B_x B_y \\ \rho v_x v_z - B_x B_z \\ (e + p + |\mathbf{B}|^2/2)v_x - B_x(\mathbf{v} \cdot \mathbf{B}) \end{pmatrix}. \quad (12.11)$$

Even though the right-hand side terms of Eq.(12.8) look very similar to second-order approximations of derivatives, this relation is in fact exact.

This formulation has the advantage of conserving the left-hand side quantities of Eqs.(12.1)–(12.3) up to machine precision, since what exits one cell enters a neighboring one and vice-versa. Mass density, momentum, and total energy are inherently conserved, which means that all numerical dissipation results in the rise of internal energy, i.e., a heating of the fluid in the adiabatic case.

12.2.4 Constrained Transport

In MHD simulations, it is very important to maintain the solenoidality of the magnetic field ($\nabla \cdot \mathbf{B} = 0$); otherwise, unphysical effects such as enhanced transport in the direction orthogonal to the local magnetic field [14] and loss of momentum and energy conservation [15] would occur. In fact, if the magnetic field is evolved

in time as a cell average, like the hydrodynamical variables, $|\nabla \cdot \mathbf{B}|$ typically grows in time.

Several ways exist which address this issue. There are some well-known techniques. First is the Helmholtz-Hodge projection [15] which exactly eliminates any dilatational component of the magnetic field but does not ensure that the projected field is physically consistent with the dynamical state of the plasma. This leads to a non-negligible generation of divergence-free but spurious small-scale magnetic fluctuations [7]. In contrast, the two other popular approaches, namely, the 8-wave multiplier method [45] and the generalized Lagrange multiplier technique [21], do not annihilate magnetic field divergences but let the plasma flow advect the erroneous field components preventing their spatial accumulation. In this context, numerical diffusion is seen as a beneficial effect as it dissipates the advected small-scale errors in the magnetic field over time. Here, we focus on the constrained transport (hereafter CT) technique [22]. It preserves magnetic field solenoidality up to machine precision in a way analogous to the cell-average conservation in finite-volume schemes. Refinements of this technique have been developed [2, 3].

In the CT approach, the magnetic field components, B_x , B_y , and B_z , are not defined as volume averages like the hydrodynamical variables but as area averages over the faces of a grid cell that are orthogonal to the \mathbf{x} –, \mathbf{y} –, and \mathbf{z} – direction, respectively (see Fig. 12.1), for example:

$$(B_x^A)_{i\pm 1/2, j, k} = \frac{1}{\Delta y \Delta z} \int_{A_{j,k}^x} B_x(x_{i\pm 1/2}, y, z) dy dz, \quad (12.12)$$

and similarly for B_y^A and B_z^A . Such a staggered discretization is also known as a Yee lattice. The face area $A_{j,k}^x$ has been defined in Sect. 12.2.3. Applying Stoke's theorem to Eq. (12.5) leads to the exact relations:

$$\partial_t (B_x^A)_{i\pm 1/2, j, k} = -\frac{D_y((E_z^L)_{i\pm 1/2, j, k})}{\Delta y} + \frac{D_z((E_y^L)_{i\pm 1/2, j, k})}{\Delta z}, \quad (12.13)$$

$$\partial_t (B_y^A)_{i, j\pm 1/2, k} = \frac{D_x((E_z^L)_{i, j\pm 1/2, k})}{\Delta x} - \frac{D_z((E_x^L)_{i, j\pm 1/2, k})}{\Delta z}, \quad (12.14)$$

$$\partial_t (B_z^A)_{i, j, k\pm 1/2} = -\frac{D_x((E_y^L)_{i, j, k\pm 1/2})}{\Delta x} + \frac{D_y((E_x^L)_{i, j, k\pm 1/2})}{\Delta y}, \quad (12.15)$$

where E^L is the line-averaged electric field (with $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$), as shown in Fig. 12.1. For example:

$$(E_z^L)_{i\pm 1/2, j+1/2, k} = \int_{z=z_k-\Delta z/2}^{z_k+\Delta z/2} E_z(x_{i\pm 1/2}, y_{j+1/2}, z) dz. \quad (12.16)$$

The second-order central approximation of $\nabla \cdot \mathbf{B}$ is then:

$$\nabla \cdot \mathbf{B} \approx \frac{D_x((B_x^A)_{i,j,k})}{\Delta x} + \frac{D_y((B_y^A)_{i,j,k})}{\Delta y} + \frac{D_z((B_z^A)_{i,j,k})}{\Delta z} \quad (12.17)$$

Using Eqs. (12.13)–(12.15), it is straightforward to show that this approximation of $\partial_t(\nabla \cdot \mathbf{B})$ is conserved, cf. [65]. Thus, if the magnetic field is initially solenoidal, then its solenoidality is preserved up to machine precision as the field is evolved in time.

12.3 Practical Computation of the Fluxes

The computation of the numerical fluxes over the faces of a grid cell is a fundamental operation in the finite-volume framework. To this end, a reconstruction step has to be performed, based on the assumption that within a grid cell all physical variables have a polynomial representation up to a given order of the cell extension, e.g., Δx . The reconstructed polynomial has to explicitly accommodate to discontinuities occurring at the grid cell boundaries. Especially for higher-order reconstruction, where multiple neighboring cells have to be taken into account, this becomes important and particularly challenging if the solution itself can feature physical discontinuities like, e.g., propagating shock fronts in high-Mach-number flows.

12.3.1 Central Weighted Essentially Non-oscillatory Reconstruction

There exist a lot of possibilities in order to compute the area-averaged values $\mathbf{U}_{i\pm 1/2,j,k}^A$ from the volume-averaged values. For a $(2n+1)$ th order procedure, the most straightforward one is to find for each quantity $q \in \mathbf{u}$ the unique polynomial $P_{i,j,k}^{opt,q}$ of degree at most $2n$ which cell averages on each of the cells of the $(2n+1)$ long stencil $\{x_{i-n}, x_{i-n+1}, \dots, x_i, x_{i+1}, \dots, x_{i+n}\}$ coincide with q 's cell average Q :

$$\frac{1}{\Delta x} \int_{x_{i+m-1/2}}^{x_{i+m+1/2}} P_{i,j,k}^{opt,q}(x) dx = \int_{\Omega_{i+m,j,k}} q dx dy dz = Q_{i+m,j,k}, \forall m \in [-n..n] \quad (12.18)$$

Then, the $\left(P_{i,j,k}^{opt,q}(x_{i\pm 1/2})\right)_{q \in \mathbf{u}}$ build a $(2n+1)$ th-order approximation of $\mathbf{U}_{i\pm 1/2,j,k}^A$. However, using such a polynomial over a discontinuity would be problematic. As shown in Fig. 12.2, such a reconstruction is oscillatory and gives large over- and undershoots. If the reconstructed value is the density ρ , for example, this would even give rise to an unphysical state with a negative density at $x = \frac{\Delta x}{2}$.

In order to tackle this problem, some methods use slope limiters which restrict the local variation of the solution to finite values by introducing numerical dissipation.

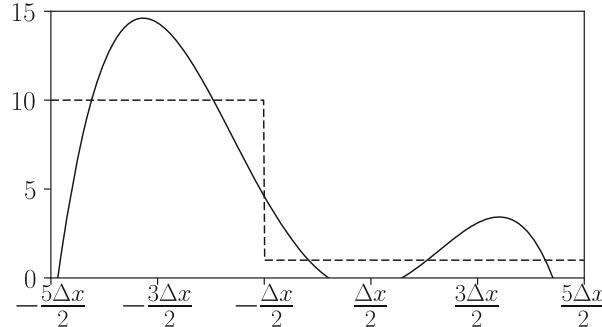


Fig. 12.2 Simple polynomial reconstruction over a discontinuity. The dashed line is the function to be reconstructed and the solid line is the unique polynomial of degree 4 in which cell averages verify Eq. (12.18). The grid size Δx is unity

Examples of such methods include the total variation diminishing (TVD) limiter of van Leer [60], a second-order procedure used, for example, in the Nirvana code [65]; the third-order Piecewise Parabolic Method [18] used, for example, in the Athena code [51]; the fifth-order MP5 scheme [52]; etc. Another class of reconstruction methods comprises the weighted essentially non-oscillatory (WENO) schemes, which we will present more in detail here. These schemes are improvements of the essentially non-oscillatory (ENO) schemes [27]. The first WENO scheme, which was third-order accurate, was designed in 1994 [36]. Since then, very high-order schemes, at least up to order 17, have been derived [10, 23]. We present here a refinement of the standard fifth-order WENO [31] used, for example, in the Pluto code [40], namely, a fourth-order central weighted non-oscillatory (CWENO) scheme [35]. For more literature about WENO schemes, the interested reader can see [49, 50]. A comparison between several solvers can be found in [33].

The idea of this CWENO scheme is the following: instead of using all the cells of a stencil at once to find P^{opt} , one divides this stencil into $n + 1$ sub-stencils. For example, for $n = 2$, the 5-cell wide stencil $\{x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}\}$ is divided in three 3-cell wide stencils containing the central cell $L = \{x_{i-2}, x_{i-1}, x_i\}$ on the left, $C = \{x_{i-1}, x_i, x_{i+1}\}$ in the center, and $R = \{x_i, x_{i+1}, x_{i+2}\}$ on the right. Then, let P_L , P_C , and P_R be the unique polynomials of degree at most 2 in which cell averages coincide with the cell averages of q on the stencils L , C , and R , respectively.¹ If q is smooth over these stencils, then each polynomial P_L , P_C , and P_R gives a third-order approximation of q on the domain $[x_{i-1/2}, x_{i+1/2}]$. One can find three positive coefficients c_L , c_C , and c_R with $c_L + c_C + c_R = 1$ such that:

$$P_{i,j,k}^{opt}(x_i \pm \Delta x/2) = c_L P_L(x_i \pm \Delta x/2) + c_C P_C(x_i \pm \Delta x/2) + c_R P_R(x_i \pm \Delta x/2). \quad (12.19)$$

¹ For the sake of simplicity, we drop the superscript q in the following, even though the polynomials are of course different for each variable.

The value of these coefficients, also called optimal weights, is $c_L = c_R = \frac{1}{6}$, $c_C = \frac{2}{3}$ for the fourth-order CWENO scheme [35].

For non-smooth problems, the idea is to replace these weights c_L , c_C , and c_R by other positive weights w_L , w_C , and w_R (with $w_L + w_C + w_R = 1$):

$$R_{i,j,k}(x_i \pm \Delta x/2) = w_L P_L(x_i \pm \Delta x/2) + w_C P_C(x_i \pm \Delta x/2) + w_R P_R(x_i \pm \Delta x/2), \quad (12.20)$$

with the weights such that in smooth regions $(w_m) = (c_m)$, $m \in \{L, C, R\}$ so as to have the reconstruction polynomial $R = P^{opt}$, providing high order of accuracy, but in regions with strong discontinuities, the polynomial(s) in which the discontinuity is present is(are) associated with a vanishing weight. For example, in Fig. 12.2, only P_R would have a non-negligible weight. This can be done through:

$$w_m = \frac{\alpha_m}{\sum \alpha_j} \quad \text{with} \quad \alpha_m = \frac{c_m}{(\epsilon + IS_m)^p}, \quad (12.21)$$

where $m \in \{L, C, R\}$; $p = 2$ as in [31, 35]; ϵ is a small positive number taken in order to avoid that the denominator becomes zero, typically $\epsilon = 10^{-6}$; and IS_m corresponds to a measure of the smoothness of the polynomial P_m :

$$IS_m = \sum_{l=1}^n \int_{x_{i-1/2}}^{x_{i+1/2}} \Delta x^{2l-1} \left(\frac{\partial^l P_m}{\partial x^l} \right)^2 dx. \quad (12.22)$$

This smoothness indicator is a measure of the total variation of the polynomial, considering as well the higher-order variations, while the Δx^{2l-1} term is there to remove the Δx terms coming from the polynomial's derivatives. All algebra done, this gives, with the subscripts (j, k) implicitly assumed:

$$IS_L = \frac{13}{12}(Q_{i-2} - 2Q_{i-1} + Q_i)^2 + \frac{1}{4}(Q_{i-2} - 4Q_{i-1} + 3Q_i)^2, \quad (12.23)$$

$$IS_C = \frac{13}{12}(Q_{i-1} - 2Q_i + Q_{i+1})^2 + \frac{1}{4}(Q_{i-1} - Q_{i+1})^2, \quad (12.24)$$

$$IS_R = \frac{13}{12}(Q_i - 2Q_{i+1} + Q_{i+2})^2 + \frac{1}{4}(3Q_i - 4Q_{i+1} + Q_{i+2})^2. \quad (12.25)$$

And the reconstructed values:

$$\begin{aligned} Q_{i+1/2}^{A,W} &= R_i(x_i + \Delta x/2), \\ &= \frac{1}{6}[w_L(2Q_{i-2} - 7Q_{i-1} + 11Q_i) + w_C(-Q_{i-1} + 5Q_i + 2Q_{i+1}) \\ &\quad + w_R(2Q_i + 5Q_{i+1} - Q_{i+2})], \end{aligned} \quad (12.26)$$

$$Q_{i-1/2}^{A,E} = R_i(x_i - \Delta x/2), \quad (12.27)$$

$$\begin{aligned}
&= \frac{1}{6} [w_L(-Q_{i-2} + 5Q_{i-1} + 2Q_i) + w_C(2Q_{i-1} + 5Q_i - Q_{i+1}) \\
&\quad + w_R(11Q_i - 7Q_{i+1} + 2Q_{i+2})].
\end{aligned}$$

Here the superscripts W and E stand, respectively, for “West” and “East,” which correspond to states obtained for $x \rightarrow x_{i+1/2}^-$ and $x \rightarrow x_{i-1/2}^+$, respectively. Please note that other choices for the smoothness indicators (see, e.g., [13, 64]) and/or the weights (see, e.g., [29]) are possible, which can reduce the dissipation, improve the accuracy near smooth extrema, and/or provide a better convergence to steady states.

When considering a system of equations, it is furthermore important that all dependent variables “feel” the discontinuities and shocks at the same locations. If each variable is reconstructed independently, oscillations can still occur. In order to avoid them, it is advisable to use so-called Global Smoothness Indicators (GSI), by taking an appropriate combination of the individual smoothness indicators of each variable and using this combination to compute the weights w_m which are common for all variables. For example, for a pure hydrodynamical problem (with $\mathbf{B} = \mathbf{0}$), it has been shown that using the smoothness indicator of the density alone for all variables is a good choice [35], and for MHD problems using a normalized average of the smoothness indicators from the density and the magnetic field components gives good results [1, 62]. An illustration of this fact is also shown in Sect. 12.6.3.1. The CWENO reconstruction procedure can hence be computed in practice through the following steps:

1. use Eqs. (12.23)–(12.25) to compute individual smoothness indicators for a well-chosen ensemble of variables,
2. combine them appropriately to compute global smoothness indicators (GSI_m), $m \in \{L, C, R\}$,
3. compute the weights using Eq. (12.21) with the (GSI_m),
4. apply these common weights in Eqs. (12.26)–(12.27) to reconstruct all the variables.

12.3.2 Reconstruction of the Magnetic Field Components

As mentioned in Sect. 12.2.4, in the CT approach, each magnetic field component is known primarily as an area average normal to its respective direction (see Fig. 12.1). In order to determine the fluxes, however, each component has to be estimated on the other faces as well. In order to do this, one can first compute the volume-averaged magnetic field inside the cells and deduce the components on any other face through the reconstruction method described above in Sect. 12.3.1. The easiest way to determine this volume average is through a polynomial interpolation, that is,

for a $2n$ th order method, one can find the unique polynomial $P_{B,A}$ of degree at most $2n - 1$ which verifies:

$$P_{B,A}(x_{i+m-1/2}) = (B_x^A)_{i+m-1/2,j,k}, \forall m \in [-n+1..n] \quad (12.28)$$

Then, the integral $\frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} P_{B,A}(x) dx$ gives a $2n$ th order approximation of $\frac{1}{\Delta x \Delta y \Delta z} \int_{\Omega_{i,j,k}} B_x dx$. All derivations done, for a fourth-order scheme, this gives:

$$\begin{aligned} \frac{1}{\Delta x \Delta y \Delta z} \int_{\Omega_{i,j,k}} B_x dx &= \frac{1}{24} \left[-(B_x^A)_{i-3/2,j,k} - (B_x^A)_{i+3/2,j,k} \right. \\ &\quad \left. + 13 \left((B_x^A)_{i-1/2,j,k} + (B_x^A)_{i+1/2,j,k} \right) \right]. \end{aligned} \quad (12.29)$$

The same method can of course be applied to obtain the volume averages of the other magnetic field components. The use of a simple interpolation is possible here since the area-averaged magnetic field components are continuous along their respective directions [37]. In order to enhance the stability of the scheme, one could, however, similar to the CWENO volume-to-area average reconstructions, use a non-oscillatory reconstruction, as in Ref. [8].

When implementing such a scheme in practice, please note that the B_m component of the magnetic field does not have to be reconstructed along the m direction since it is known as a primary data. This also means that $(B_x^{A,E})_{i\pm1/2,j,k} = (B_x^{A,W})_{i\pm1/2,j,k} = (B_x^A)_{i\pm1/2,j,k}$ and similar for the other directions, which is consistent with the continuity of B_m along direction m .

12.3.3 Solving the Riemann Problem

Through the reconstruction procedure, two states for (\mathbf{U}, \mathbf{B}) are determined at each interface between cells: a “West” and an “East” one. At each interface, we are thus facing a so-called Riemann problem² and want to estimate the fluxes through this interface. We do not detail here how such problems are solved in practice, but the interested reader can see, for example, [58] for details. Numerically, several Riemann solvers with different levels of complexity have been proposed (see, e.g., Ref. [34, 42, 46, 47]). We present here the simplest one, the so-called Rusanov flux [47] (also known as “local Lax-Friedrichs” or LLF [37]):

$$\mathbf{f}_{LLF}^x = \frac{1}{2} \left(\mathbf{f}^x \left(\mathbf{U}^E, \mathbf{B}^E \right) + \mathbf{f}^x \left(\mathbf{U}^W, \mathbf{B}^W \right) \right) - \frac{a^x}{2} \left(\mathbf{U}^E - \mathbf{U}^W \right), \quad (12.30)$$

² A Riemann problem is an initial value problem with a system in a certain constant state on one side of an impermeable interface and a different constant state on the other side, presenting hence a discontinuity. The central question is the system’s temporal evolution after removal of the interface.

where \mathbf{f}^x is the physical flux in the x -direction (see Eq.(12.11)) and a^x is the maximum local speed of propagation of information in the system in the x -direction, which corresponds in the case of the MHD equations to the fast-magnetosonic wave:

$$a^x = \max \left((|v_x| + c_f^x)^W, (|v_x| + c_f^x)^E \right), \quad (12.31)$$

with c_f^x the magneto-sonic speed:

$$c_f^x = \sqrt{\frac{1}{2} \left((c_s^2 + c_A^2) + \sqrt{(c_s^2 + c_A^2)^2 - 4c_s^2 \frac{B_x^2}{\rho}} \right)}, \quad (12.32)$$

where $c_s = (\gamma p / \rho)^{1/2}$ and $c_A = (|\mathbf{B}|^2 / \rho)^{1/2}$ are the sound speed and the Alfvén speed, respectively. The Rusanov flux has the advantage of being very simple and computationally inexpensive. However, it is way more dissipative than other more accurate Riemann solvers, such as the Roe [46] or the HLLD solver [42]. Please note: if one wants to have a scheme of order higher than 2, one cannot simply use Eq.(12.30) with the computed area averages. The next section addresses this issue.

12.3.4 Passage Through Point Values

As can be seen from Eq.(12.11), quantities such as $\iint \rho v_x^2$ need to be known in order to compute the fluxes. From the reconstruction step (Sects. 12.3.1 and 12.3.2), the area averages $\iint \rho$ and $\iint \rho v_x$ are known. Mathematically, however, $\iint \rho v_x^2 \neq \frac{(\iint \rho v_x)^2}{\iint \rho}$. A way to walk around this difficulty is to compute point values from the area averages, since, for point values, $\rho v_x^2 = \frac{(\rho v_x)^2}{\rho}$. The main idea to perform such a transformation comes from a Taylor expansion and can be found in [17, 39]. For the sake of simplicity, we derive first the formula for a one-dimensional problem:

$$Q_i = \frac{1}{\Delta x} \int_{\epsilon=-\Delta x/2}^{\Delta x/2} q(x_i + \epsilon) d\epsilon, \quad (12.33)$$

$$= \frac{1}{\Delta x} \int \left(q(x_i) + \epsilon q'(x_i) + \frac{\epsilon^2}{2!} q''(x_i) + \frac{\epsilon^3}{3!} q'''(x_i) + O(\epsilon^4) \right) d\epsilon, \quad (12.34)$$

$$= q(x_i) + \frac{\Delta x^2}{24} q''(x_i) + O(\Delta x^4). \quad (12.35)$$

Through this formula, one can also see that $Q_i = q_i + O(\Delta x^2)$. Thus, identifying the spatial average with a point value in the middle of the considered domain results in an error of second order. This is why we need to consider the next term, $\frac{\Delta x^2}{24}q''(x_i)$, if we want to have a fourth-order scheme. It can be shown that a second-order approximation of $q''(x_i)$ is [17]:

$$q''(x_i) = \frac{q_{i-1} - 2q_i + q_{i+1}}{\Delta x^2} + O(\Delta x^2) = \frac{Q_{i-1} - 2Q_i + Q_{i+1}}{\Delta x^2} + O(\Delta x^2), \quad (12.36)$$

where the first equality contains point values and the second cell averages. Please note that, except in this special case, one cannot in general simply replace point values by cell averages in the formulas.

The Taylor expansion of Eq.(12.34) can be generalized to more than one dimension. For a 2D fourth-order area-average-to-point transformation, this leads to the following expression:

$$\begin{aligned} q_{i\pm 1/2,j,k} &= Q_{i\pm 1/2,j,k} - \frac{Q_{i\pm 1/2,j+1,k} - 2Q_{i\pm 1/2,j,k} + Q_{i\pm 1/2,j-1,k}}{24} \\ &\quad - \frac{Q_{i\pm 1/2,j,k+1} - 2Q_{i\pm 1/2,j,k} + Q_{i\pm 1/2,j,k-1}}{24} + O(\Delta x^4 + \Delta y^4 + \Delta z^4). \end{aligned} \quad (12.37)$$

For schemes of order higher than 4, cross-derivatives (e.g., $\partial_x^2 \partial_y^2$) need to be considered.

The computed point values are used in order to determine point-valued fluxes through Eq.(12.30), which are then transformed using a point-to-area-average transformation. Equations (12.36) and (12.35) lead to:

$$\begin{aligned} Q_{i\pm 1/2,j,k} &= q_{i\pm 1/2,j,k} + \frac{q_{i\pm 1/2,j+1,k} - 2q_{i\pm 1/2,j,k} + q_{i\pm 1/2,j-1,k}}{24} \\ &\quad + \frac{q_{i\pm 1/2,j,k+1} - 2q_{i\pm 1/2,j,k} + q_{i\pm 1/2,j,k-1}}{24} + O(\Delta x^4 + \Delta y^4 + \Delta z^4). \end{aligned} \quad (12.38)$$

Plugging \mathbf{f}_{LLF}^x (Eq. (12.30)) there gives a higher-order approximation to the area-averaged flux $\mathbf{F}^{A,x}$. This area-averaged flux is used in Eq. (12.8) in order to evolve the cell averages of the hydrodynamical quantities in time (see Sect. 12.4).

Once again, please note that the fact that Eqs. (12.38) and (12.37) are looking very similar to each other is a special case, valid only because of the similarity in Eq. (12.36). This similarity is lost for schemes of order higher than 4.

12.3.5 Electric Fluxes on the Edges

In order to evaluate the right-hand side for the magnetic field evolution, we have to take care of a two-dimensional Riemann problem. Indeed, formally, Eq. (12.5) can be reformulated in a conservative form:

$$\partial_t \mathbf{B} = -\nabla \cdot \begin{pmatrix} 0 & -E_z & E_y \\ E_z & 0 & -E_x \\ -E_y & E_x & 0 \end{pmatrix}. \quad (12.39)$$

This means that the electric field plays the role of a flux. Following Sect. 12.3.4, electric field point values can be computed from the velocity and magnetic field point values and then transformed into area averages, two per interface between two cells, for example, $\mathbf{E}_{i+1/2,j,k}^{A,W}$ and $\mathbf{E}_{i+1/2,j,k}^{A,E}$ at the interface between cells (i, j, k) and $(i + 1, j, k)$. Then, the same CWENO procedure as the one described in Sect. 12.3.1 can be applied to these area averages in the two remaining directions in order to deduce line averages: in our example we would obtain for a reconstruction in the y -direction the four states $\mathbf{E}_{i+1/2,j+1/2,k}^{L,SW}$, $\mathbf{E}_{i+1/2,j+1/2,k}^{L,SE}$, $\mathbf{E}_{i+1/2,j-1/2,k}^{L,NW}$, and $\mathbf{E}_{i+1/2,j-1/2,k}^{L,NE}$ where the superscripts S and N stand, respectively, for “South” ($y \rightarrow y_{j+1/2}^-$) and “North” ($y \rightarrow y_{j-1/2}^+$). For these area-to-line CWENO reconstructions, one can take for Eqs. (12.26)–(12.27) the weights computed by taking the mean of the *GSI* in the reconstruction direction of the two cells surrounding the area (see Sect. 12.3.1). In the end, at each edge surrounded by four faces, we have four different states for the electric field (see Fig. 12.3), which are computed by taking the mean of the two possible reconstructed values (e.g., one can obtain $\mathbf{E}_{i+1/2,j+1/2,k}^{L,SW}$ by (a) reconstructing $\mathbf{E}_{i+1/2,j,k}^{A,W}$ along y or (b) reconstructing $\mathbf{E}_{i,j+1/2,k}^{A,S}$ along x). These

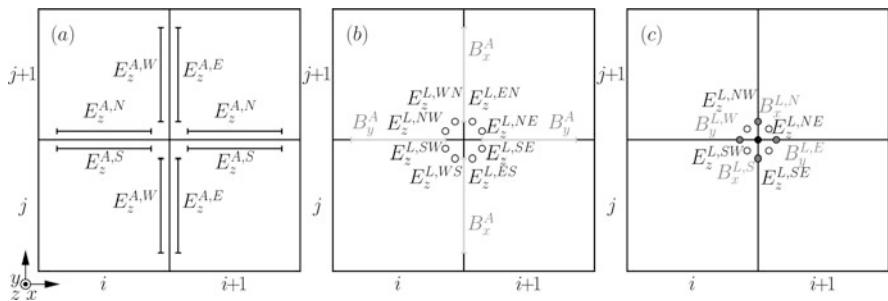


Fig. 12.3 Illustration of the computation of $(E_z^L)_{i+1/2,j+1/2,k}$ (black circle in subgraph (c)): (a) CWENO line-average reconstruction of the area-averaged electric field, (b) CWENO line-average reconstruction of the area-averaged magnetic field and taking the mean of the two possibilities for the electric field in each quadrant, (c) deducing the electric flux, thanks to the multidimensional Riemann solver

four states are then plugged in a 2D Riemann solver. We present here the LLF variant of the multidimensional Riemann solver of reference [4]:

$$(E_z^L) = \frac{1}{4} \left[(E_z^L)^{SE} + (E_z^L)^{SW} + (E_z^L)^{NE} + (E_z^L)^{NW} \right] \quad (12.40)$$

$$+ \frac{S^a}{2} \left((B_y^L)^E - (B_y^L)^W \right) - \frac{S^a}{2} \left((B_x^L)^N - (B_x^L)^S \right),$$

where the quantities with a superscript L are computed through area-to-line average reconstructions and are indicated in Fig. 12.3 and S^a is the maximum speed of propagation of information, estimated as:

$$S_{i+1/2,j+1/2,k}^a = \max(a_{i+1/2,j,k}^x, a_{i+1/2,j+1,k}^x, a_{i,j+1/2,k}^y, a_{i+1,j+1/2,k}^y). \quad (12.41)$$

This choice for evaluating S^a is done for reasons of efficiency, since the local maximum speeds $a^{x,y}$ have already been computed in order to solve the 1D Riemann problems (see Sect. 12.3.3), as mentioned in [38]. Several refinements of this solver are available in the literature: in Ref. [4], a solver with four different speeds can be found and further improvements have been made in Ref. [9].

12.3.6 Summary: The Complete Procedure to Determine the RHS

The numerical method to compute the right-hand sides of Eqs. (12.8) and (12.13)–(12.15) can be summarized as follows:

1. Compute cell averages for the magnetic field components (Sect. 12.3.2).
2. For each dimension $n \in \{x, y, z\}$:
 - (a) reconstruct area averages of all quantities but B_n along n (Sects. 12.3.1 and 12.3.2),
 - (b) transform the area averages to point values (Sect. 12.3.4),
 - (c) deduce point-valued fluxes (Sect. 12.3.3) and electric field (Sect. 12.3.5),
 - (d) transform the point-valued fluxes and electric field to area averages (Sect. 12.3.4),
 - (e) deduce one term of Eq. (12.8).
3. For each dimension $n \in \{x, y, z\}$:
 - (a) reconstruct line averages of the \mathbf{B} and \mathbf{E} components normal to n (Sect. 12.3.5),
 - (b) deduce through Eq. (12.41) two terms in Eqs. (12.13)–(12.15).

The dimension-by-dimension approach presented here is computationally more efficient than an actual multidimensional CWENO reconstruction which requires bi (resp. tri)-quadratic polynomials in 2D (resp. 3D)[61].

12.4 Time Integration

Numerous explicit time integration methods have been developed with different mathematical properties making them more appropriate in different contexts. When dealing with problems with strong discontinuities and shocks, the so-called strong stability preserving Runge–Kutta (SSPRK) methods are advantageous since they avoid additional oscillations generated by the time integration step [25, 26]. Furthermore, their linear stability region is typically greater as compared to standard Runge–Kutta schemes, allowing larger timesteps and hence a large gain in efficiency.

The numerical tests presented in this chapter are done using the ten-stage fourth-order SSPRK method presented in pseudocode 3 of Ref. [32], which is repeated for the sake of completeness in Algorithm 1. In this pseudocode, Δt is the timestep, w represents the whole state of the system (the hydrodynamical variables as well as the magnetic field) initially at instant t but at the last line at instant $t + \Delta t$, $F(k_1)$ is the right-hand side computed from the state k_1 , and s the stage number. Three registers are needed: k_1 , k_2 and one to store $F(k_1)$. In order to obtain a stable solution with correct results, the timestep Δt cannot be arbitrarily large but is restricted by the Courant-Friedrichs-Lowy (CFL) stability criterion [19]. We present here a formulation of this criterion that can be found in [57], for example:

$$\Delta t \leq C_{CFL} \min_{i,j,k} \left(\frac{\Delta x}{a_{i,j,k}^x}, \frac{\Delta y}{a_{i,j,k}^y}, \frac{\Delta z}{a_{i,j,k}^z} \right), \quad (12.42)$$

where C_{CFL} is a constant called the Courant number, $a^n (n \in \{x, y, z\})$ is the maximum speed of propagation of information in the n -direction, which in our case corresponds to fast magnetosonic waves and is expressed in Eq. (12.31). In the case of the ten-stage method used here, even though a high number of stages are performed, its stability region is very large, allowing a high Courant constant $C_{CFL} = 1.95$ for the 2D and $C_{CFL} = 1.5$ for the 3D systems we consider in this chapter.

Algorithm 1 10 stages SSPRK4 method from [32]

```

 $k_1 \leftarrow w$ 
for  $s = 1 : 5$  do
     $k_1 \leftarrow k_1 + \frac{\Delta t}{6} F(k_1)$ 
end for
 $k_2 \leftarrow \frac{1}{25} w + \frac{9}{25} k_1$ 
 $k_1 \leftarrow 15k_2 - 5k_1$ 
for  $s = 6 : 9$  do
     $k_1 \leftarrow k_1 + \frac{\Delta t}{6} F(k_1)$ 
end for
 $w \leftarrow k_2 + \frac{3}{5} k_1 + \frac{\Delta t}{10} F(k_1)$ 

```

In order to spare computing time, it is advisable to allow the timestep to vary. At each iteration, the maximum allowed timestep verifying Eq.(12.42) is computed during the first stage of the SSPRK method and kept the same for the remaining stages, until the next iteration.

12.5 Strong Shocks and Negative Pressure/Density

Even when using a non-oscillatory reconstruction and a SSPRK time-integration method, negative densities or pressure may still appear when dealing with strong discontinuities or shocks. The increase of the stiffness of the reconstruction polynomials together with the reduction of numerical dissipation which are both consequences of the increase of the algorithm's precision makes high-order schemes particularly vulnerable to this issue. Indeed, even though numerical dissipation causes amplitude errors in the solution, which tend to smooth out small-scale structures, it has the positive side effect of smoothing out unphysical reconstruction variations as well. Because of this, lower-order schemes deliver less precise solutions but are more robust when compared to higher order schemes. A negative density or pressure can arise after the reconstruction step (Sect. 12.3.1), area-to-point transformation (Sect. 12.3.4), or time integration (Sect. 12.4).

A possibility to deal with this issue is to use a *fallback* approach, which consists of using lower-order reconstruction procedures in the vicinity of strong shocks, to smooth out the discontinuity and avoid additional oscillations.

First of all, one can use an a posteriori fallback approach: if after the reconstruction procedure (Sect. 12.3.1) a negative density or pressure is obtained, then a lower-order reconstruction is performed locally, such as a second-order TVD limiter [60], or in the extreme case Godunov's scheme $\mathbf{U}_{i\pm 1/2,j,k}^A = \mathbf{U}_{i,j,k}$. Similarly, if the area-to-point transformation would give an unphysical result, it can be switched off locally: the area averages are identified with the point values.

However, this may not be enough for a strongly shocked problem. An idea then is to use an a priori fallback approach, also called flattening [5, 18], which would reconstruct at lower order even though the higher-order methods still give a physical state such as to avoid the appearance of unphysical states in the next timesteps. In practice, assume a high-order reconstruction method *HO* and a low-order reconstruction method *LO*, which give the reconstructed states $(\mathbf{U}^A, \mathbf{B}^A)_{i\pm 1/2,j,k}^{HO}$ and $(\mathbf{U}^A, \mathbf{B}^A)_{i\pm 1/2,j,k}^{LO}$, respectively. Then, the reconstructed state at the interfaces $(i \pm 1/2, j, k)$ is given by:

$$(\mathbf{U}^A, \mathbf{B}^A)_{i\pm 1/2,j,k} = w_{i,j,k}^f (\mathbf{U}^A, \mathbf{B}^A)_{i\pm 1/2,j,k}^{HO} + (1 - w_{i,j,k}^f) (\mathbf{U}^A, \mathbf{B}^A)_{i\pm 1/2,j,k}^{LO}, \quad (12.43)$$

with a weight $w_{i,j,k}^f \in [0, 1]$ called the “flattener,” equal to 1 in smooth regions and going to 0 as the considered region is more and more shocked. In a similar way, this flattening also occurs for the area-to-point transformation in case the lower-order

method is first- or second-order accurate:

$$\begin{aligned} q_{i \pm 1/2, j, k} &\approx Q_{i \pm 1/2, j, k} - w_{i \pm 1/2, j, k}^f \frac{Q_{i \pm 1/2, j+1, k} - 2Q_{i \pm 1/2, j, k} + Q_{i \pm 1/2, j-1, k}}{24} \\ &\quad - w_{i \pm 1/2, j, k}^f \frac{Q_{i \pm 1/2, j, k+1} - 2Q_{i \pm 1/2, j, k} + Q_{i \pm 1/2, j, k-1}}{24}, \end{aligned} \quad (12.44)$$

and similar for the point-to-area transformation. There are several possible choices for the flattener: in the XTROEM-FV code [43], the Jameson indicator [30] is used. Another flattener, well suited for the structure of the MHD equations, can be found in [5], which also presents refinements in order to include cells that are not affected by the shock yet but will shortly be. We present, however, here one inspired by Ref. [18] which uses the pressure gradient, that is, we define first, for a reconstruction in the x -direction, for example, the shock indicator:

$$s_{i, j, k}^x = \frac{|\tilde{p}_{i+1, j, k} - \tilde{p}_{i-1, j, k}|}{\tilde{p}_{i, j, k}}, \quad (12.45)$$

with $\tilde{p}_{i, j, k}$ an estimate of the pressure in cell (i, j, k) computed by considering the cell averages as if they were point values. Then, we choose two thresholds $\tau^{HO} < \tau^{LO}$ and define the flattener w_f as:

$$w_{i, j, k}^f = \begin{cases} 1, & \text{if } s_{i, j, k}^x < \tau^{HO}, \\ 1 - \frac{s_{i, j, k}^x - \tau^{HO}}{\tau^{LO} - \tau^{HO}}, & \text{if } \tau^{HO} \leq s_{i, j, k}^x \leq \tau^{LO}, \\ 0, & \text{if } \tau^{LO} < s_{i, j, k}^x. \end{cases} \quad (12.46)$$

The choice of the thresholds τ^{HO} and τ^{LO} is dependent both on the problem and the reconstruction method. If the lower-order method is second order, a reasonable first try is $\tau^{HO}, \tau^{LO} \lesssim 4$. Indeed, if one would use naively a linear approximation, without limiters, to reconstruct the pressure at the cell interfaces, $p_{i \pm 1/2, j, k} = \tilde{p}_{i, j, k} \pm \frac{\tilde{p}_{i+1, j, k} - \tilde{p}_{i-1, j, k}}{2\Delta x} \cdot \frac{\Delta x}{2}$, then a value of $s_{i, j, k}^x = 4$ would already give a zero pressure on one of the interfaces. The distance between τ^{LO} and τ^{HO} is there to provide a smooth transition from high to low order. In practice, this rough upper boundary may be too high for highly shocked problem and may need to be decreased severely depending on the dynamical state of the system.

For the flattener at the area-to-point-value transformations in Eq. (12.44), one can take a function of the w^f used for reconstruction of the neighboring cells. Adding a superscript $n \in \{x, y, z\}$ corresponding to the reconstruction direction, we decided to use:

$$w_{i+1/2, j, k}^f = \min \left(w_{i, j, k}^{f,y}, w_{i+1, j, k}^{f,y}, w_{i, j, k}^{f,z}, w_{i+1, j, k}^{f,z} \right) \quad (12.47)$$

This method is of course generalizable to more than two reconstruction methods and area↔point value transformations. Another possibility to deal with strong discontinuities may be WENO schemes which automatically adapt their order in their vicinity [6], or the use of provably positivity preserving schemes [63], at the probable expense of more computational resources.

12.6 Numerical Tests

In the case of the numerical scheme presented here, the following aspects have to be verified:

1. the scheme should converge toward an exact solution in continuous space and time at the proper rate in the asymptotic limit of $\Delta x, \Delta t \rightarrow 0, \Delta t/\Delta x$ finite. This is the object of Sect. 12.6.1.
2. The solenoidality of the magnetic field should be preserved up to machine precision: in all the tests performed here, $\nabla \cdot \mathbf{B}$ is close to machine precision and does not grow over time. Since the $\nabla \cdot \mathbf{B}$ is *preserved* in the CT approach, we would like to point out that the initialization is very important in this respect. In practice, it is a good idea to use analytical expressions for the area averages of the magnetic field components and not point values in the middle of the faces, for example, even when using a second-order scheme.
3. The scheme should be robust, that is, be able to handle correctly problems with shocks and discontinuities. In particular, it should not give unphysical values such as a negative pressure or density. This is the object of Sect. 12.6.3.

In all the tests, we take $\gamma = \frac{5}{3}$, which corresponds physically to a mono-atomic gas with three degrees of freedom, and $C_{CFL} = 1.95$ for 1D and 2D problems and $C_{CFL} = 1.5$ for 3D problems, unless mentioned otherwise. The boundary conditions are periodic in all tests, apart for the Brio-Wu Riemann problem.

12.6.1 Verification of the Scheme's Order

For a smooth problem, the error between the solution given by the numerical scheme and the exact solution is expected to diminish asymptotically with increasing grid resolution at a well-defined rate corresponding to the scheme's order. A possibility to perform a convergence test is to take a smooth problem with a periodic motion and compare the solution after a period $t = T$ to the initial conditions at $t = 0$. When the exact solution is not known, the order of convergence can be computed by comparing to a very high resolution run which is then taken as the reference (see the convergence test presented in Sect. 12.6.3.2). There are several possibilities to

compute the discretization error. We present here the same one as in [41] which uses the L_1 norm:

$$\delta\mathbf{W} = \frac{1}{N_x N_y N_z} \sum_{i,j,k} |\mathbf{W}_{i,j,k}(t = T) - \mathbf{W}_{i,j,k}(t = 0)| \quad (12.48)$$

with $\mathbf{W} = (\mathbf{U}, \mathbf{B})$ so that $\delta\mathbf{W}$ is the error vector for all the eight MHD variables. The number of grid points in each direction is N_x , N_y , and N_z . We define furthermore $\delta W_{mean} = \frac{1}{8} \sum_i \delta\mathbf{W}_i$ the mean of these errors. In order to measure the convergence, the same numerical test is repeated for several resolutions r_1, r_2, \dots, r_N and the experimental order of convergence EOC is given by:

$$EOC_m = \frac{|\log(\delta W_{mean}(r_m)) - \log(\delta W_{mean}(r_{m-1}))|}{|\log(r_m) - \log(r_{m-1})|} \quad (12.49)$$

and should be asymptotically equal to the scheme's order n .

12.6.2 Smooth Problems

Smooth problems are useful to determine the order of convergence of a numerical scheme. We present here a circularly polarized Alfvén wave and a 3D MHD vortex. Furthermore, the fourth-order scheme, denoted in the following by CWENO4, is compared with a second-order scheme denoted TVD2 in order to show some advantages of using higher-order numerics. For the second-order scheme, we make use of the TVD limiter of van Leer [60] instead of the fourth-order CWENO reconstruction:

$$Q_{i+1/2}^{A,W} = Q_i + \frac{\max[(Q_{i+1} - Q_i)(Q_i - Q_{i-1}), 0]}{Q_{i+1} - Q_{i-1}}, \quad (12.50)$$

$$Q_{i-1/2}^{A,E} = Q_i - \frac{\max[(Q_{i+1} - Q_i)(Q_i - Q_{i-1}), 0]}{Q_{i+1} - Q_{i-1}}. \quad (12.51)$$

In addition, the area averages \leftrightarrow point value transformations are switched off (i.e., the transformations of Sect. 12.3.4 are identity functions). We maintain, however, the fourth-order magnetic field area-to-volume interpolation as well as the fourth-order time integration so as to be consistent with the a priori fallback approach for the shocked problems (see Sect. 12.6.3.3). In the tests performed here, the present CWENO4 scheme is about 75% more expensive than the TVD2 one. This estimate is of course system, optimization, and implementation dependent. The moderate increase of computational effort comes together with a rise in precision that permits to use lower numerical resolution as compared to lower-order techniques for a result of similar quality.

For these smooth problems, the fallback approach is not needed at all and has been deactivated. To show the need of the passage through point values, we consider

a CWENO4A scheme, which is the CWENO4 one when ignoring the passage through point values, plugging directly area averages in the Riemann solver.

12.6.2.1 Circularly Polarized Alfvén Wave

Alfvén waves are exact and smooth, inherently linear solutions of the nonlinear MHD equations which are widely used to check the accuracy of a scheme [41, 43, 51, 59]. We use here the same initial conditions as the circularly polarized Alfvén wave in [65]: $\rho = 1$, $\mathbf{v} = (-\frac{A}{\sqrt{2}} \sin(2\pi(x+y)), \frac{A}{\sqrt{2}} \sin(2\pi(x+y)), A \cos(2\pi(x+y)))$, $\mathbf{B} = (\frac{B_0}{\sqrt{2}} + \frac{A}{\sqrt{2}} \sin(2\pi(x+y)), \frac{B_0}{\sqrt{2}} - \frac{A}{\sqrt{2}} \sin(2\pi(x+y)), -A \cos(2\pi(x+y)))$ and $p = 0.1$, with the wave amplitude $A = 0.1$ and the mean magnetic field $B_0 = \sqrt{2}$ in the computational domain $(x, y) \in [0, 1]^2$.

The convergence of errors is measured by comparing the solution after one period (at $t = 0.5$, since the wave is propagating at the Alfvén speed $\sqrt{\frac{B_0^2}{\rho}} = \sqrt{2}$ along the diagonal) to the initial conditions. Table 12.1 shows the results for the TVD2, CWENO4, and CWENO4A schemes. The relative energy dissipation E_{loss} is measured by $E_{loss} = \frac{E_T(t=0) - E_T(t=T)}{E_T(t=0)}$ with E_T the total energy (kinetic+magnetic) but removing the mean fields. We can see that for such a linear solution, even when omitting the passage through point values, the fourth-order scheme gives fourth-order convergence. This is the reason why nonlinear tests, like the one in the next section, have to be performed as well in order to check the convergence order properly. Figure 12.4 shows as well some effects of the higher numerical dissipation of the lower-order schemes on the wave amplitude after 100 periods, at $t = 50$. In order to achieve an as low dissipation as the CWENO4 scheme at resolution 128^2 , the TVD2 scheme needs to go to resolutions higher than 1024^2 (Table 12.1), leading to a way higher computational cost, about a factor 300.

Table 12.1 Circularly polarized Alfvén wave: convergence of errors, EOC, and numerical dissipation for different schemes after one period

Resolution		32^2	64^2	128^2	256^2	512^2	1024^2
TVD2	δU_{mean}	2.46×10^{-3}	7.03×10^{-4}	1.76×10^{-4}	4.18×10^{-5}	9.88×10^{-6}	2.31×10^{-6}
	EOC	—	1.81	2.00	2.07	2.08	2.09
	E_{loss}	4.20×10^{-2}	5.14×10^{-3}	6.20×10^{-4}	7.53×10^{-5}	9.21×10^{-6}	1.13×10^{-6}
CWENO4	δU_{mean}	1.50×10^{-5}	4.86×10^{-7}	1.53×10^{-8}	5.08×10^{-10}	1.87×10^{-11}	9.95×10^{-13}
	EOC	—	4.95	4.99	4.91	4.76	4.23
	E_{loss}	7.01×10^{-4}	2.46×10^{-5}	7.75×10^{-7}	2.27×10^{-8}	5.91×10^{-10}	1.37×10^{-11}
CWENO4A	δU_{mean}	1.51×10^{-5}	4.87×10^{-7}	1.59×10^{-8}	5.79×10^{-10}	2.49×10^{-11}	1.42×10^{-12}
	EOC	—	4.95	4.94	4.78	4.54	4.14
	E_{loss}	7.01×10^{-4}	2.45×10^{-5}	7.75×10^{-7}	2.27×10^{-8}	5.91×10^{-10}	1.41×10^{-11}

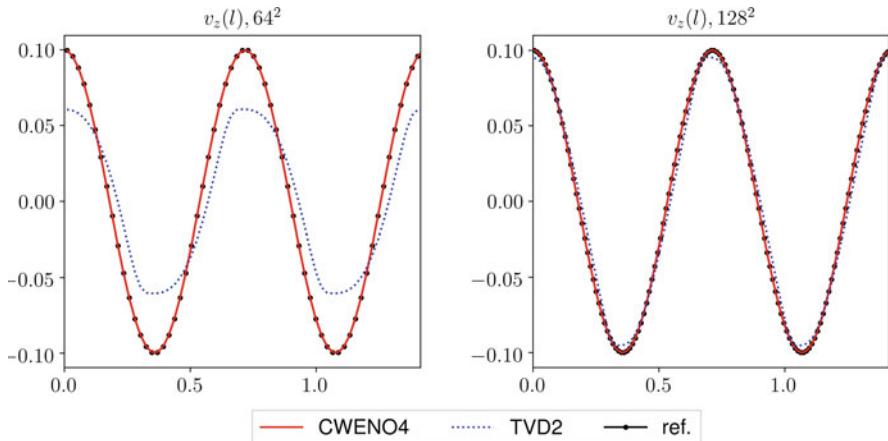


Fig. 12.4 Circularly polarized Alfvén wave: cuts of v_z along the main diagonal ($l = \sqrt{x^2 + y^2} \in [0, \sqrt{2}]$) after 100 periods at two different resolutions (64^2 and 128^2). The CWENO4 solution (solid red) is extremely close to the reference solution at $t = 0$ (black, circle markers), whereas the TVD2 solution (dotted blue) exhibits significant amplitude and phase errors

12.6.2.2 3D MHD Vortex

The MHD vortex problem was first introduced in 2D in [2]. The initial conditions consist of a magnetized vortex structure in force equilibrium that is advected by a velocity field. We present here a 3D extension of this test proposed in [41]. The initial conditions are given by $\rho = 1$, $\mathbf{v} = (1 - y\kappa \exp[q(1 - r^2)], 1 + x\kappa \exp[q(1 - r^2)], 2)$, $\mathbf{B} = (-y\mu \exp[q(1 - r^2)], x\mu \exp[q(1 - r^2)], 0)$ and $p = 1 + \frac{1}{4q}[\mu^2(1 - 2q(r^2 - z^2)) - \kappa^2\rho] \exp[2q(1 - r^2)]$ with $r^2 = x^2 + y^2 + z^2$, $\kappa = \mu = 1/(2\pi)$ and $q = 1$ on the computational domain $(x, y, z) \in [-5, 5]^3$. Here as well, the solution after one period ($t = 10$ since the advection speed is 1 along \mathbf{x} and \mathbf{y}) is compared to the initial fields. Table 12.2 shows that for this nonlinear test problem, omitting the passage through point values leads to a convergence order of 2, which is expected (see Eq. (12.33)). This table also shows that in order to obtain a result's quality as the one at resolution 128^3 for the CWENO4 scheme, the TVD2 scheme needs to be performed at resolution 512^3 , leading to a computational cost increase of about a factor 150. Figure 12.5 presents some effects of the numerical dissipation on the vortex' shape and amplitude.

12.6.3 Shocked Problems

In order to test how well numerical schemes handle shocks, a lot of classical tests have been designed. We present in this chapter the 1D Brio-Wu Riemann problem and the 2D Orszag-Tang vortex. Through these two tests, we aim as well at underlining two distinct aspects of the CWENO4 scheme, respectively: first,

Table 12.2 3D MHD vortex problem: convergence of errors, EOC, and numerical dissipation after one period

Resolution		32^3	64^3	128^3	256^3	512^3	1024^3
TVD2	δU_{mean}	8.85×10^{-4}	4.33×10^{-4}	1.35×10^{-4}	3.92×10^{-5}	1.02×10^{-5}	2.59×10^{-6}
	EOC	—	1.03	1.68	1.79	1.94	1.98
	E_{loss}	0.92	0.58	0.15	2.02×10^{-2}	2.39×10^{-3}	2.85×10^{-4}
CWENO4	δU_{mean}	4.22×10^{-4}	4.35×10^{-5}	2.12×10^{-6}	8.86×10^{-8}	4.40×10^{-9}	2.56×10^{-10}
	EOC	—	3.28	4.36	4.58	4.33	4.11
	E_{loss}	0.58	6.36×10^{-2}	2.23×10^{-3}	7.00×10^{-5}	2.04×10^{-6}	5.43×10^{-8}
CWENO4A	δU_{mean}	4.21×10^{-4}	4.37×10^{-5}	2.30×10^{-6}	1.64×10^{-7}	2.98×10^{-8}	7.26×10^{-9}
	EOC	—	3.27	4.25	3.81	2.46	2.04
	E_{loss}	0.58	6.36×10^{-2}	2.29×10^{-3}	7.02×10^{-5}	2.09×10^{-6}	6.74×10^{-8}

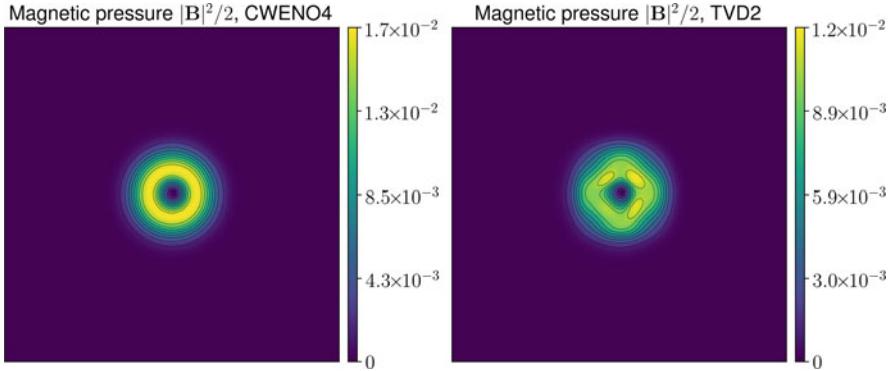


Fig. 12.5 3D MHD vortex problem: slices of the magnetic pressure at $z \approx 0.04$, resolution 128^3 after one period for the CWENO4 and TVD2 schemes

the importance of using global smoothness indicators (as compared to the use of different ones for each variable) and, second, that convergence studies need a smooth flow in order to reveal the order of a numerical scheme. Other classical problems to test the robustness of a numerical scheme include the MHD rotor problem (e.g., [11, 51, 59, 65]), MHD blast waves (e.g., [37, 43, 51, 65]), the current sheet problem (e.g., [28, 43]), the cloud-shock interaction (e.g., [20, 43, 59, 65]), etc. How a numerical scheme almost identical to the present one performs on some of these tests is shown in [62].

Finally, we show an example where the use of an a priori fallback approach is necessary, namely, a supersonic decaying turbulence experiment, and illustrate some advantages of higher-order numerics over second-order schemes in terms of resolution of small-scale structures.

12.6.3.1 1D Brio-Wu Riemann Problem

One-dimensional Riemann problems serve as standard benchmarks in computational MHD and allow to evaluate the robustness of the numerical method as

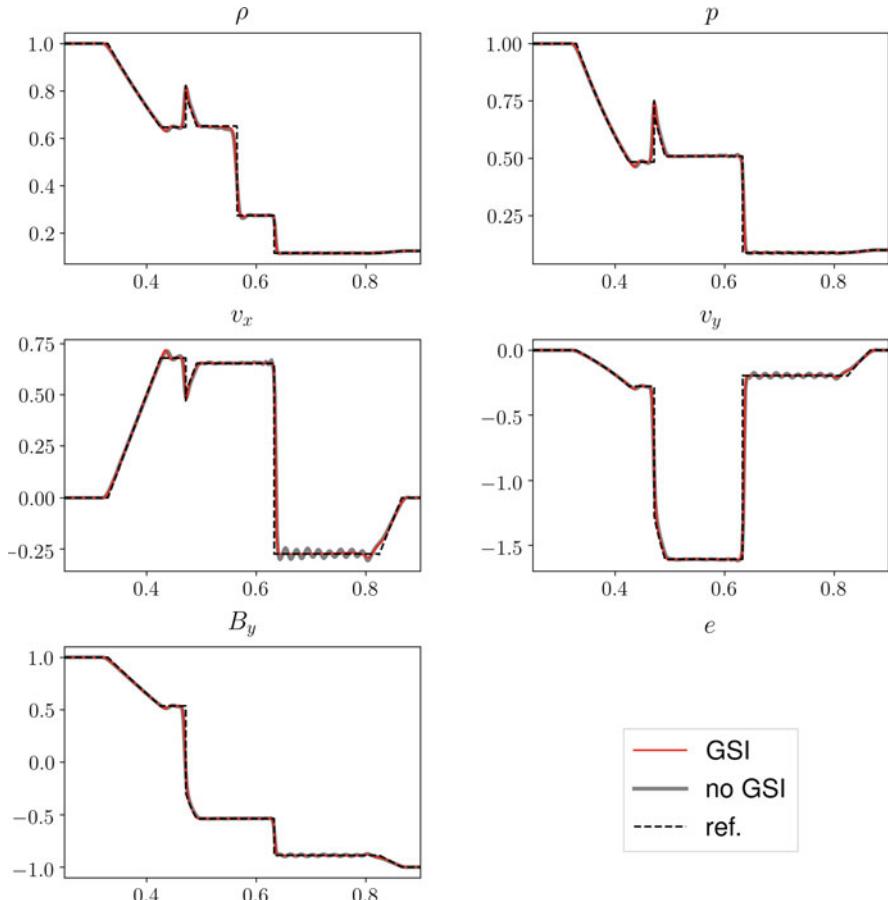


Fig. 12.6 Solution of the Brio-Wu Riemann problem at $t = 0.1$, CWENO4 scheme. Red and gray: using respectively global smoothness indicators (GSI) and individual ones, resolution 512. Black, dashed: reference solution obtained with GSI at resolution 65536. Five waves separated by constant states can be seen, from the left to the right: a fast rarefaction wave and a slow compound wave both moving to the left, a contact discontinuity, a slow shock, and a fast rarefaction wave, all three moving to the right

well as how good it can resolve discontinuities, shocks, and appropriate types of waves. In [48], a lot of Riemann problems are proposed for testing several physical aspects. We present here the results obtained for the classical Brio-Wu shock tube test [16, 48]. The initial conditions for $(\rho, v_x, v_y, v_z, p, B_x, B_y, B_z)$ consist of the constant states $(1, 0, 0, 0, 1, 0.75, 1, 0)$ on the left ($x \leq 0.5$) and $(0.125, 0, 0, 0, 0.1, 0.75, -1, 0)$ on the right ($x > 0.5$) of the domain $x \in [0, 1]$.

In this test, open boundary conditions are used. Figure 12.6 shows the solution at time $t = 0.1$ and provides a comparison between the use of individual/global

smoothness indicators. As mentioned at the end of Sect. 12.3.1, the use of global smoothness indicators reduces strongly the amount of oscillations.

12.6.3.2 Orszag–Tang Vortex

The Orszag–Tang vortex was first studied in the framework of incompressible flows [44] and has become a standard test to check the robustness of compressible MHD solvers ([41, 43, 51, 59, 65] among many others). From smooth initial conditions, many shock structures emerge and develop turbulent dynamics. The initial conditions are given in the computational domain $(x, y) \in [0, 1]^2$ by $\rho = \gamma^2$, $\mathbf{v} = (-\sin(2\pi y), \sin(2\pi x), 0)$, $\mathbf{B} = (-\sin(2\pi y), \sin(4\pi x), 0)$ and $p = \gamma$.

Table 12.3 shows a convergence test done at two instants: $t = 0.1$ when the solution is still smooth and at $t = 0.5$ with fully developed shocks (Fig. 12.7). Since the exact solution is not known, the point values located at the cell centers of the lowest resolution run (32^2) are compared to a high resolution outcome (323^2) considered as the reference. We can see that even though the convergence order is consistent with a fourth-order scheme at $t = 0.1$, it goes to 1 at later times when the convergence errors are dominated by the shocks, which is consistent with Godunov’s theorem [24].

12.6.3.3 Decaying Supersonic MHD Turbulence

As an example of application where the fallback approach is needed, we present here a 3D decaying turbulence experiment at resolution 512^3 . In order to avoid the heating of the plasma when kinetic and magnetic energy are dissipated, we use here the isothermal equation of state (see Sect. 12.2.1 regarding implications for the solver) with $c_s = 0.1$ on a box $(x, y, z) \in [0, 1]^3$. The initial velocity and magnetic field components are generated in Fourier space by giving each mode with wavenumber \mathbf{k} an amplitude proportional to $\exp(-k^2/(2k_0^2))$ with $k_0 = 4$, similar to [12]. The fields are generated in such a way that kinetic helicity, magnetic helicity, and cross helicity are initially 0 and the initial velocity field is solenoidal. The fields are then transformed into configuration space and normalized so as to give a root-mean-square Mach number of $M = 2$ (with a minimum local value of 1.4×10^{-3} and a maximum of 5.6) and an initial ratio of magnetic to kinetic energy of 1. The density is initially $\rho = 1$ everywhere.

Without a priori fallback approach, the CWENO4 scheme gives a negative density shortly after about 0.2 large eddy turnover time. The large eddy turnover

Table 12.3 Orszag–Tang vortex: convergence of errors and EOC at different instants

Resolution		32^2	96^2	288^2	544^2	800^2	1056^2
$t = 0.1$	δU_{mean}	1.56×10^{-2}	3.96×10^{-4}	3.84×10^{-6}	2.55×10^{-7}	4.96×10^{-8}	1.54×10^{-8}
	EOC	–	3.34	4.22	4.27	4.24	4.21
$t = 0.5$	δU_{mean}	2.43×10^{-1}	8.79×10^{-2}	2.77×10^{-2}	1.36×10^{-2}	8.44×10^{-3}	5.75×10^{-3}
	EOC	–	0.91	1.04	1.11	1.24	1.39

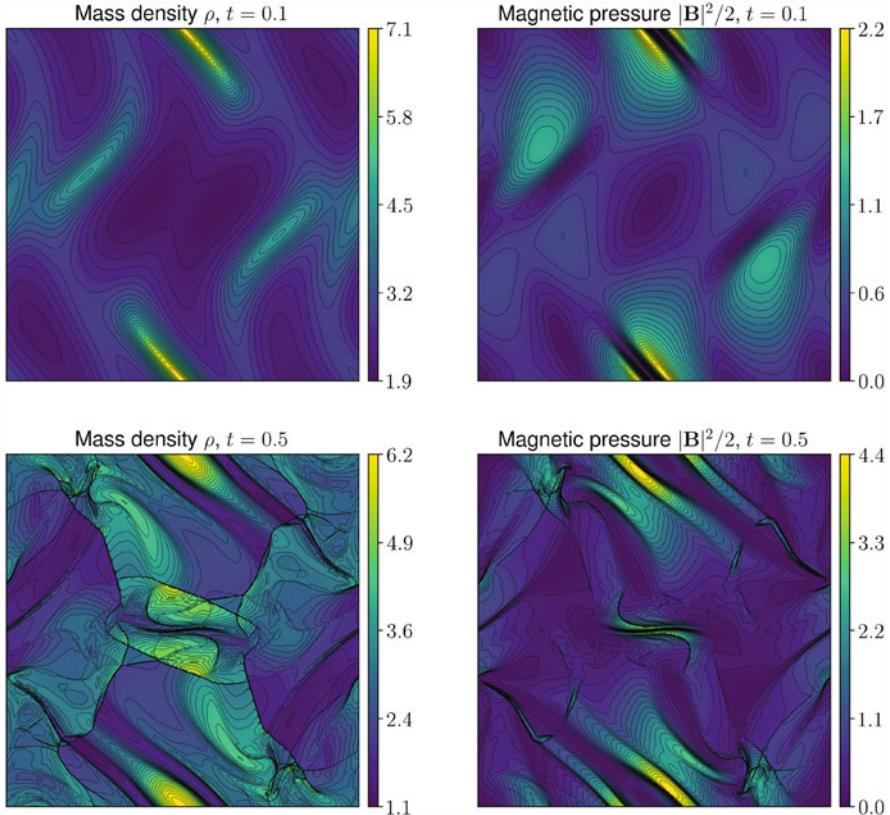


Fig. 12.7 Orszag–Tang vortex plots, resolution 1056^2 , CWENO4 scheme, at different instants

time t_E is evaluated through $t_E = \frac{L_I}{\sqrt{E(0)}}$ with $E(0) = \frac{1}{2}\rho(0)\mathbf{v}(0)^2 + \mathbf{B}(0)^2$ and $L_I = \frac{\int_k k^{-1}E(k)dk}{\int_k E(k)dk}$ its integral scale. Using the TVD2 scheme as the lower-order method and the thresholds $\tau^{HO} = 1$, $\tau^{LO} = 2$ permits to run the simulation without giving negative densities at any time. Only few reconstructions are performed at lower order: the number of reconstructions using a flattener w^f strictly lower than 1 (see Sect. 12.5) is maximal at around 0.55 large eddy turnover time and concerns about 6/1000 reconstructions. We denote this scheme by CWENO4FB in the following and compare it with TVD2.

At the very beginning of the decay (less than $0.1t_E$), magnetic energy is rapidly transformed into kinetic energy, accelerating the fluid. The velocity field also causes energy to be stored in a potential form, as density fluctuations, which account for about 10–20% of the kinetic energy during the run. On a timescale of the order of t_E , fluctuations over the full range of resolved scales are excited by nonlinear energy transfer from the largest scales of the flow, and a state of fully developed turbulence

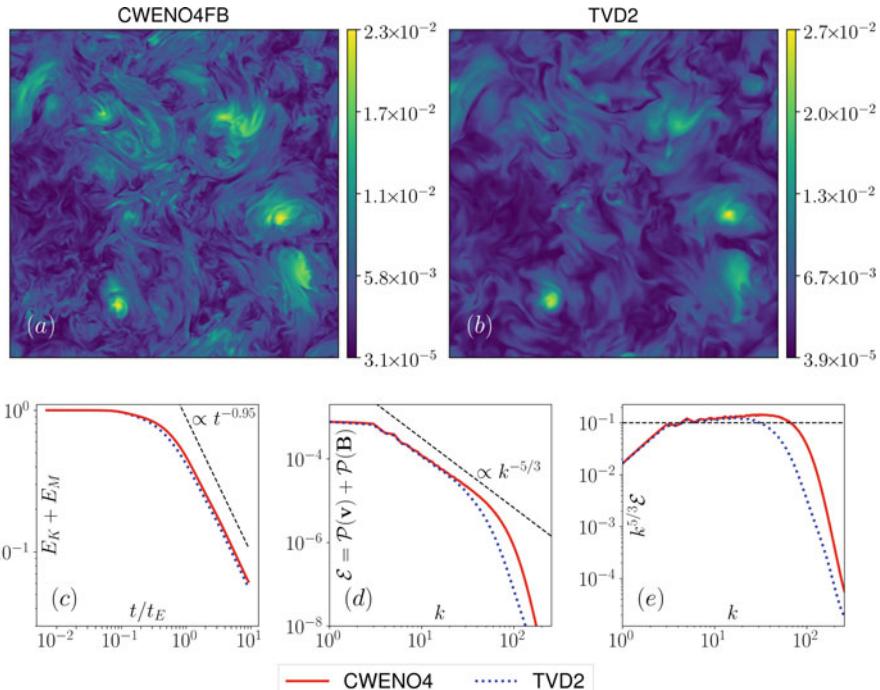


Fig. 12.8 Decaying supersonic 512^3 MHD turbulence plots comparing the CWENO4FB and the TVD2 schemes. **(a–b)** Slices of kinetic plus magnetic energy at $t \approx 4.20t_E$, with t_E the large eddy turnover time, CWENO4FB scheme. **(c)** Log-log plot of the kinetic plus magnetic energy as a function of time (in units of t_E), normalized by its initial value. The dashed black line is a guide to the eye and corresponds to the power law $t^{-0.95}$. **(d)** Specific kinetic $\mathcal{P}(\mathbf{v})$ plus magnetic energy $\mathcal{P}(\mathbf{B})$ spectra at $t \approx 4.20t_E$, the dashed black line is a guide to the eye with power law $k^{-5/3}$. **(e)** Same spectra $\mathcal{E} = \mathcal{P}(\mathbf{v}) + \mathcal{P}(\mathbf{B})$ but compensated with $k^{5/3}$

emerges. The regime of decaying turbulence starts: the energy decays, following a power law $E(t) \sim t^{-\alpha}$ with an exponent close to 1 (Fig. 12.8c), similar to the observations made in the nonhelical incompressible case [12]. During the run, the maximum local Mach number peaked at around 7.4; shortly after $t = 0.45t_E$, the maximum local density peaked at about 51.6, the minimum at around 0.01.

Figure 12.8 shows comparisons between the CWENO4FB and TVD2 schemes: we can see that the CWENO4FB scheme presents clearly more small-scale structures at $t \approx 4.20t_E$ which are smeared out by the second-order scheme. This is also visible when considering the specific kinetic plus magnetic energy Fourier spectra: a Kolmogorov-like inertial range with an exponent close to $-5/3$ is visible for both schemes, but obviously longer for the CWENO4FB scheme. However, when compensating these spectra with $k^{5/3}$, the exponent of the CWENO4FB scheme looks slightly flatter than the one of the TVD2 scheme. This well-known effect is probably due to the hyperdiffusive characteristics of the numerical scheme and can be alleviated by adding explicit physical diffusion terms for viscosity and resistivity.

12.7 Final Remarks

In the recent decade, significant progress has been made with regard to higher-order finite-volume algorithms. This contribution outlines the necessary efforts to go beyond the standard and widespread implementation of second-order simulation algorithms for ideal magnetohydrodynamics in the conservative finite-volume framework. The gains in precision, detail, and numerical consistency with ideal Euler dynamics are substantial. The numerical cost remains well acceptable and even decreases dramatically as higher-order can offer the same quality as lower-order ones at strongly reduced numerical resolution. As the elimination of significant amounts of numerical dissipation reveals new difficulties that have been covered up by numerical diffusion before, strategies of regularizing the solution in the presence of strong shocks are proposed. Due to the strong reduction of numerical dissipation, schemes that do not strictly preserve the magnetic field solenoidality but rely on advection and diffusion might not be as efficient in higher-order simulations. Fortunately, the constrained transport approach represents a viable solution whose implementation is described here, as well.

As an example of application, a decaying supersonic turbulence experiment has been presented, for which small-scale structures are clearly better resolved using higher-order numerics. This illustrates that higher-order schemes provide more reliable simulations of astrophysical flows, since turbulence very often plays a crucial role there. In particular, a variant of the present algorithm has been used to study the inverse transfer of magnetic helicity in supersonic isothermal MHD turbulence [53–56]. Magnetic helicity is an ideal invariant of the MHD equations which is found in astrophysical flows and have the tendency to form large-scale structures from small-scale fluctuations, making it a potential candidate to explain the presence of large-scale magnetic fields, for example, in the interstellar medium. As shown in [53], a lower-order scheme dissipates significantly more magnetic helicity, making the study of its dynamics less reliable, whereas clear scaling laws can be observed with a higher-order scheme.

In summary, at the current state of numerical developments, it seems to be highly appropriate and rewarding to consider moving existing simulation codes to higher-order schemes.

The contributions of P.-S. Verma and O. Henze to the present work are cordially acknowledged. JMT acknowledges support by the Berlin International Graduate School in Model and Simulation based Research (BIMoS).

References

1. J. Balbas, E. Tadmor, Nonoscillatory central schemes for one- and two-dimensional magnetohydrodynamics equations. II high-order semidiscrete schemes. *Soc. Ind. Appl. Math. J. Sci. Comput.* **28**(2), 533–560 (2006)
2. D.S. Balsara, Second-order-accurate schemes for magnetohydrodynamics with divergence-free reconstruction. *Astrophys. J. Suppl. Series* **151**, 149–184 (2004)

3. D.S. Balsara, Divergence-free reconstruction of magnetic fields and WENO schemes for magnetohydrodynamics. *J. Comput. Phys.* **228**, 5040–5056 (2009)
4. D.S. Balsara, Multidimensional HLLE Riemann solver: application to Euler and magnetohydrodynamic flows. *J. Comput. Phys.* **229**, 1970–1993 (2010)
5. D.S. Balsara, Self-adjusting, positivity preserving high order schemes for hydrodynamics and magnetohydrodynamics. *J. Comput. Phys.* **231**, 7504–7517 (2012)
6. D.S. Balsara, S. Garain, C.-W. Shu, An efficient class of WENO schemes with adaptive order. *J. Comput. Phys.* **326**, 780–804 (2016)
7. D.S. Balsara, J. Kim, A comparison between divergence-cleaning and staggered-mesh formulations for numerical magnetohydrodynamics. *Astrophys. J.* **602**, 1079–1090 (2004)
8. D.S. Balsara, C. Meyer, M. Dumbser, H. Du, Z. Xu, Efficient implementation of ADER schemes for Euler and magnetohydrodynamical flows on structured meshes—speed comparisons with Runge–Kutta methods. *J. Comput. Phys.* **235**, 934–969 (2013)
9. D.S. Balsara, B. Nkonga, Multidimensional Riemann problem with self-similar internal structure—part III—a multidimensional analogue of the HLLI Riemann solver for conservative hyperbolic systems. *J. Comput. Phys.* **346**, 25–48 (2017)
10. D.S. Balsara, C.-W. Shu, Monotonicity preserving weighted essentially non-oscillatory schemes with increasingly high order of accuracy. *J. Comput. Phys.* **160**, 405–452 (2000)
11. D.S. Balsara, D.S. Spicer, A staggered mesh algorithm using high order Godunov fluxes to ensure solenoidal magnetic fields in magnetohydrodynamic simulations. *J. Comput. Phys.* **149**, 270–292 (1999)
12. D. Biskamp, W.-C. Müller, Scaling properties of three-dimensional isotropic magnetohydrodynamic turbulence. *Phys. Plasmas* **7**, 4889–4900 (2000)
13. R. Borges, M. Carmona, B. Costa, W.S. Don, An improved weighted essentially non-oscillatory scheme for hyperbolic conservation laws. *J. Comput. Phys.* **227**, 3191–3211 (2008)
14. J. Brackbill, Fluid modeling of magnetized plasmas. *Space Sci. Rev.* **42**, 153–167 (1985)
15. J.U. Brackbill, D.C. Barnes, The effect of nonzero $\nabla \cdot b$ on the numerical solution of the magnetohydrodynamic equations. *J. Comput. Phys.* **35**, 426–430 (1980)
16. M. Brio, C.C. Wu, An upwind differencing scheme for the equations of ideal magnetohydrodynamics. *J. Comput. Phys.* **75**, 400–422 (1988)
17. P. Buchmüller, C. Helzel, Improved accuracy of high-order WENO finite volume methods on Cartesian grids. *J. Sci. Comput.* **61**, 343–368 (2014)
18. P. Colella, P.R. Woodward, The piecewise parabolic method (PPM) for gas-dynamical simulations. *J. Comput. Phys.* **54**, 174–201 (1984)
19. R. Courant, K. Friedrichs, H. Lewy, Über die partiellen Differenzengleichungen der mathematischen Physik. *Mathematische Annalen* **100**, 32–74 (1928). English: IBM J. Res. Dev. **11** (1967)
20. W. Dai, P.R. Woodward, Numerical simulations for radiation hydrodynamics. I. Diffusion limit. *J. Comput. Phys.* **142**, 182–207 (1998)
21. A. Dedner, F. Kemm, D. Kröner, C.-D. Munz, T. Schnizer, M. Wesenberg, Hyperbolic divergence cleaning for the MHD equations. *J. Comput. Phys.* **175**, 645–673 (2002)
22. C.R. Evans, J.F. Hawley, Simulation of magnetohydrodynamic flows: a constrained transport method. *Astrophys. J.* **332**, 659–677 (1988)
23. G. Gerolymos, D. Sénéchal, I. Vallet, Very-high-order WENO schemes. *J. Comput. Phys.* **228**, 8481–8524 (2009)
24. S.K. Godunov, A difference scheme for numerical solution of discontinuous solution of hydrodynamic equations. *Matematicheskii Sbornik* **47**, 271–306 (1959). English: US Joint Publications Research Service 7226, 1969
25. S. Gottlieb, C.-W. Shu, Total variation diminishing Runge–Kutta schemes. *Math. Comput.* **67**(221), 73–85 (1998)
26. S. Gottlieb, C.-W. Shu, E. Tadmor, Strong stability-preserving high-order time discretization methods. *Soc. Ind. Appl. Math. Rev.* **51**(1), 89–112 (2001)
27. A. Harten, B. Engquist, S. Osher, S.R. Chakravarthy, Uniformly high order accurate essentially non-oscillatory schemes, III. *J. Comput. Phys.* **71**, 231–303 (1987)

28. J.F. Hawley, J.M. Stone, MOCCT: a numerical technique for astrophysical MHD. *Comput. Phys. Commun.* **89**, 127–148 (1995)
29. A.K. Henrick, T.D. Aslam, J.M. Powers, Mapped weighted essentially non-oscillatory schemes: achieving optimal order near critical points. *J. Comput. Phys.* **207**, 542–567 (2005)
30. A. Jameson, W. Schmidt, E. Turkel, Numerical solution of the Euler equations by finite volume methods using Runge–Kutta time-stepping schemes, in *AIAA 14th Fluid and Plasma Dynamic Conference* (1981)
31. G.-S. Jiang, C.-W. Shu, Efficient implementation of weighted ENO schemes. *J. Comput. Phys.* **126**, 202–228 (1996)
32. D.I. Ketcheson, Highly efficient strong stability-preserving Runge–Kutta methods with low-storage implementations. *Soc. Ind. Appl. Math. J. Sci. Comput.* **30**(4), 2113–2136 (2008)
33. A.G. Kritsuk et al., Comparing numerical methods for isothermal magnetized supersonic turbulence. *Astrophys. J.* **737**, 13 (2011)
34. A. Kurganov, S. Noelle, G. Petrova, Semidiscrete central-upwind schemes for hyperbolic conservation laws and Hamilton-Jacobi equations. *Soc. Ind. Appl. Math. J. Sci. Comput.* **23**(3), 707–740 (2001)
35. D. Levy, G. Puppo, G. Russo, Central WENO schemes for hyperbolic systems of conservation laws. *Math. Model. Numer. Anal.* **33**(3), 547–571 (1999)
36. X.-D. Liu, S. Osher, T. Chan, Weighted essentially non-oscillatory schemes. *J. Comput. Phys.* **115**, 200–212 (1994)
37. P. Londrillo, L. Del Zanna, High-order upwind schemes for multidimensional magnetohydrodynamics. *Astrophys. J.* **530**, 508–524 (2000)
38. P. Londrillo, L. Del Zanna, On the divergence-free condition in Godunov-type schemes for ideal magnetohydrodynamics: the upwind constrained transport method. *J. Comput. Phys.* **195**, 17–48 (2004)
39. P. McCorquodale, P. Colella, A high-order finite-volume method for conservation laws on locally refined grids. *Commun. Appl. Math. Comput. Sci.* **6**(1), 1–25 (2011)
40. A. Mignone, G. Bodo, S. Massaglia, T. Matsakos, O. Tesileanu, C. Zanni, A. Ferrari, Pluto: a numerical code for computational astrophysics. *Astrophys. J. Suppl. Series* **170**, 228–242 (2007)
41. A. Mignone, P. Tzeferacos, G. Bodo, High-order conservative finite difference GLM-MHD schemes for cell-centered MHD. *J. Comput. Phys.* **229**, 5896–5920 (2010)
42. T. Miyoshi, K. Kusano, A multi-state HLL approximate Riemann solver for ideal magnetohydrodynamics. *J. Comput. Phys.* **208**, 315–343 (2005)
43. J. Núñez-de la Rosa, C.-D. Munz, XTROEM-FV: A new code for computational astrophysics based on very high-order finite volume methods—I. Magnetohydrodynamics. *Monthly Notices R. Astron. Soc.* **455**, 3458–3479 (2016)
44. A. Orszag, C.M. Tang, Small-scale structure of two-dimensional magnetohydrodynamic turbulence. *J. Fluid Mech.* **90**, 129–143 (1979)
45. K.G. Powell, An approximate Riemann solver for magnetohydrodynamics (that works in more than one dimension). ICASE Report No. 94-24 (1994)
46. P.L. Roe, Approximate Riemann solvers, parameter vectors, and difference schemes. *J. Comput. Phys.* **43**, 357–372 (1981)
47. V.V. Rusanov, The calculation of the interaction of non-stationary shock waves with barriers. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki* **1**(2), 267–279, 1961 English: USSR Comput. Math. Math. Phys. **1**(2), 304–320 (1962)
48. D. Ryu, T.W. Jones, Numerical magnetohydrodynamics in astrophysics: algorithm and tests for one-dimensional flow. *Astrophys. J.* **442**, 228–258 (1995)
49. C.-W. Shu, Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws. ICASE Report No. 97-65 (1997)
50. C.-W. Shu, High order weighted essentially nonoscillatory schemes for convection dominated problems. *Soc. Ind. Appl. Math. Rev.* **51**(1), 82–126 (2009)
51. J.M. Stone, T.A. Gardiner, P. Teuben, J.F. Hawley, J.B. Simon, Athena: a new code for astrophysical MHD. *Astrophys. J. Suppl. Series* **178**, 137–177 (2008)

-
- 52. A. Suresh, H.T. Huynh, Accurate monotonicity-preserving schemes with Runge–Kutta time stepping. *J. Comput. Phys.* **136**, 83–99 (1997)
 - 53. J.-M. Teissier, *Magnetic Helicity Inverse Transfer in Isothermal Supersonic Magnetohydrodynamic Turbulence*. Ph.D. Thesis, Technische Universität Berlin, 2020
 - 54. J.-M. Teissier, W.-C. Müller, Inverse transfer of magnetic helicity in supersonic magnetohydrodynamic turbulence. *J. Phys. Conf. Series* **1623**, 012011 (2020)
 - 55. J.-M. Teissier, W.-C. Müller, Inverse transfer of magnetic helicity in direct numerical simulations of compressible isothermal turbulence: scaling laws. *J. Fluid Mech* **915**, A23 (2021)
 - 56. J.-M. Teissier, W.-C. Müller, Inverse transfer of magnetic helicity in direct numerical simulations of compressible isothermal turbulence: helical transfers. *J. Fluid Mech* **921**, A7 (2021)
 - 57. V.A. Titarev, E.F. Toro, ADER schemes for three-dimensional non-linear hyperbolic systems. *J. Comput. Phys.* **204**, 715–736 (2005)
 - 58. E.F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics* (Springer, Heidelberg, 2009)
 - 59. G. Tóth, The $\nabla \cdot b = 0$ constraint in shock-capturing magnetohydrodynamics codes. *J. Comput. Phys.* **161**, 605–652 (2000)
 - 60. B. van Leer, Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection. *J. Comput. Phys.* **23**, 276–299 (1977)
 - 61. P.S. Verma, W.-C. Müller, Higher order finite volume central schemes for multi-dimensional hyperbolic problems. *J. Sci. Comput.* **75**, 941–969 (2018)
 - 62. P.S. Verma, J.-M. Teissier, O. Henze, W.-C. Müller, Fourth order accurate finite volume CWENO scheme for astrophysical MHD problems. *Monthly Notices of R. Astron. Soc.* **482**, 416–437 (2019)
 - 63. K. Wu, C.-W. Shu, Provably positive high-order schemes for ideal magnetohydrodynamics: analysis on general meshes. *Numer. Math.* **142**, 995–1047 (2019)
 - 64. S. Zhang, C.-W. Shu, A new smoothness indicator for the WENO schemes and its effect on the convergence to steady state solutions. *J. Sci. Comput.* **31**(1/2), 273–305 (2007)
 - 65. U. Ziegler, A central-constrained transport scheme for ideal magnetohydrodynamics. *J. Comput. Phys.* **196**, 393–416 (2004)



EMAPS: An Intelligent Agent-Based Technology for Simulation of Multiscale Systems

13

Yuri Omelchenko and Homa Karimabadi

Abstract

Many systems in science and engineering can be modeled by partial differential equations (PDEs) discretized in time and space. While a variety of techniques have been developed to address multiple spatial scales frequently occurring in these systems, time integration continues to be commonly carried out by time stepping. In its simplest form, when a computational system is explicitly updated based on the smallest global timestep, this (“time-driven”) integration becomes often intractable for problems with wide ranges of temporal scales. Thus, one is forced to either suppress processes occurring on faster scales or break a single global timestep into a number of smaller, hierarchically ordered timesteps. Ideally, all elements of a computational system, composed of spatially distributed discrete quantities (e.g., cell-based fields, particles, etc.), should be updated in accordance with their own levels of useful activity (“change”). In general, this is not feasible within the time-stepping paradigm. To this end, we have developed a new simulation technology: an event-driven multi-agent planning system (EMAPS). It may be regarded as a general simulation time operating system, applicable beyond the domain of physics-based simulation. EMAPS converts a given multiscale model into a numerical system composed of interacting intelligent agents and a scheduler agent. This numerical system evolves in simulation time based on asynchronous event detection and prediction, not on time stepping. The EMAPS approach has demonstrated superior stability and accuracy properties compared to the time-stepped methods across a wide range of applications, attaining significant improvements in simulation speed.

Y. Omelchenko (✉)

Space Science Institute, Boulder, CO, USA

H. Karimabadi

Analytics Ventures, Inc., San Diego, CA, USA

(up to several orders of magnitude depending on the problem) and numerical accuracy. An EMAPS-based hybrid (kinetic ions, fluid electrons) code, HYPERS has been used to conduct some of the largest and most accurate three-dimensional simulations of the Earth’s magnetosphere to date. In this chapter we discuss the algorithmic foundation of this technology and illustrate its advantages through a number of physical examples. Particular emphasis in these examples is on hybrid simulations of space and laboratory plasmas. Potential applications of EMAPS to other domains such as climate modeling, amorphous computing, neural computation, and artificial intelligence are also discussed.

13.1 Introduction

13.1.1 Time Versus Change

Differential equations are often written as functions of continuous or discrete time. Numerical techniques for scientific and engineering applications commonly assume that these equations are properly discretized in space on a suitable spatial mesh composed of discrete control elements—cells. A mesh can be either uniform or adaptive to spatial scales. Simulating a complex system with discrete variables (e.g., cell-defined field components, Lagrangian particles, etc.) evolving on disparate timescales often presents numerical challenges.

The ongoing debate on whether time is a true dimension or a virtual construction to help us understand changes in the universe [1] directly relates to numerical simulation. Indeed, should one consider the evolution of a physical model using time as an additional dimension or should one evolve the model by following local changes instead? The former stipulates setting timesteps, which dictate update rates of the global state of the system, whereas the latter defines meaningful physical changes, which control the choice of local time increments needed for advancing model equations.

Using time as a primary factor for “driving” a numerical model, however, exposes immediate deficiencies. Time stepping is not concerned with predicting changes or detecting them “on the fly.” In its simplest case, the entire model is advanced in time with the most restrictive Courant-Friedrichs-Lowy (CFL) timestep, even when changes (if any are present in the system at all) occur at different physical rates across a global simulation domain. The single-timestep approach, however, often becomes intractable in global multiscale simulations where quantities of interest are characterized by spatially disparate rates of change. Using a small timestep for all variables may significantly slow down such simulations and in addition increase numerical diffusion. Many attempts have been made to numerically address multiple timescales, with varying degrees of success (see more on this below). Still, no complete solution has emerged on the basis of time stepping for a number of reasons. First, the CFL condition is only a linear criterion for numerical stability, not a general accuracy condition. Second, this condition does not take into account source terms in the model equations and random effects. As a result, the minimum

CFL timestep does not always guarantee numerical stability or accuracy of strongly coupled nonlinear terms, especially if they contain noisy components or reactive terms.

We have taken a different, change-based approach to simulation, and believe that a postulated local change, df should drive a necessary local update time increment, dt , not the other way around. By following this principle, we have developed an intelligent multi-agent technology, EMAPS: *Event-driven Multi-Agent Planning System*, which can be regarded as a general *simulation time operating system* concept, applicable beyond the domain of physics-based simulation. In essence, EMAPS converts a multiscale physical model into a computational system composed of interacting intelligent agents and a scheduler agent (“engine”). The engine orchestrates a Discrete-Event Simulation (DES) of the numerical system and performs global tasks. This “event-driven” system evolves in simulation time based on asynchronous event (change) detection and prediction, not on time stepping.

Figure 13.1 provides a high-level view of EMAPS by highlighting its intelligent agents, i.e., autonomous entities that act and direct their activities toward achieving their goals. A detailed discussion of the algorithm follows in Sect. 13.2, in the context of simulations described in Sects. 13.3–13.5. In EMAPS, intelligent agents are responsible for the local states of their assigned cells. The work of these agents has an added requirement of preserving global conservation laws in asynchronous local updates, as explained in Sects. 13.3 and 13.4. In particular, Sect. 13.4 describes a novel, parallel event-driven hybrid code, HYPERs. Several prospective applications of EMAPS are discussed in Sect. 13.6.

13.1.2 Time-Stepping Approaches to Asynchronous Time Integration

In this short overview we will limit ourselves to discussing only explicit time integration since implicit integration is synchronous by its nature (unless it incorporates elements of explicit techniques). To maintain numerical stability, the traditional explicit time-stepping techniques require that a simulation model is advanced in time with increments that satisfy appropriate CFL requirements. As already noted above, this methodology becomes inefficient for systems with disparate timescales. Previously, a variety of workarounds have been proposed to enable asynchronous updates within the time-stepping paradigm. For instance, an adaptive mesh refinement (AMR) technique was proposed for block-structured meshes [2]. In this technique, refined (“daughter”) meshes overlap regions covered by the coarser (“parent”) mesh so that the global mesh is made of a hierarchy of nested levels of logically rectangular patches. Each mesh patch then may be updated with an individual time increment, usually governed by the most restrictive local CFL number. Other researchers (e.g., see [3] and the references in [4]) implemented local time stepping (multi-stepping). In these methods different time increments are selected as fractions of the global timestep size in order to satisfy local stability

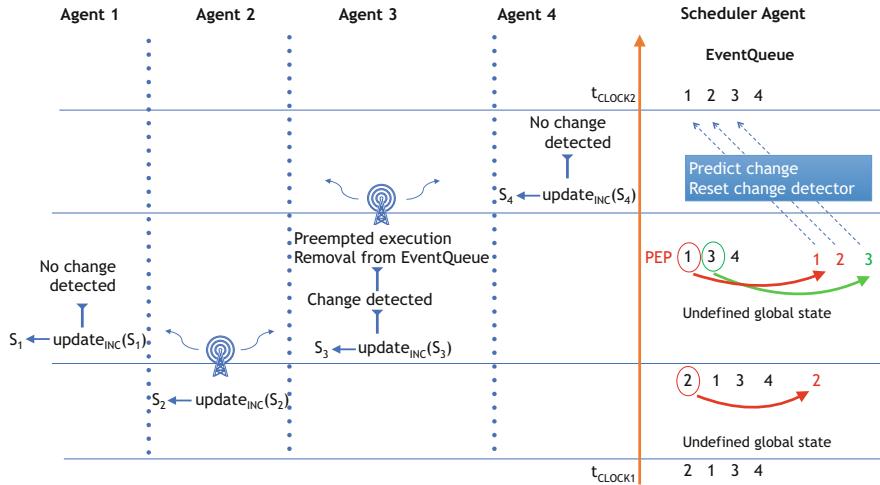


Fig. 13.1 Evolution of an intelligent multi-agent environment from simulation time, $t = t_{clock1}$ to $t = t_{clock2}$ (bottom to top). Each agent handles a local state, S_i together with its built-in change detector (“flux capacitor”) and predictor. The red and green colors refer to states processed in the scheduled order and out of order (preemptively), respectively. The “update” method updates the agent state, while the “update_{INC}” method updates both the agent state and change detector. The global system update begins when the scheduler agent (engine) processes the top event in EventQueue, which passes flow control to Agent 2. Agent 2 updates and broadcasts its state via synchronization protocols to its dependent agents, 1 and 3. These agents update their states and change detectors and further propagate the update sequence if they detect that significant changes have occurred since their last predictions. The engine collectively processes eligible (close in time) top events in the scheduled order at $t = t_{clock2}$ in the Preemptive Event Processing (PEP) mode. The new global state is defined by rescheduling all synchronized states, processed directly by the engine in EventQueue or preempted by local agents due to detected changes

conditions. In most (but not all) cases special care is taken to preserve flux continuity across cell interfaces in asynchronous updates.

Another time integration strategy has been applied to equations of nonlinear elastodynamics [5]. It is based on a discrete space-time form of Hamilton’s variational principle. This algorithm permits selection of independent timesteps at cells so that local timesteps are not necessarily integer multiples of the smallest (reference) timestep. This is achieved by organizing computational update times in a priority queue based on predictions that *are not modified* during the evolution of the system. This asynchronous approach permits updating local subsystems with frequencies dictated by their natural timescales, subject to solvability of local timesteps. More general techniques (e.g., [6] and the references therein) introduce an extra dimension—an explicitly defined *temporal mesh*. A discontinuous Galerkin solution is then constructed in time-space by adapting the duration of each element.

A common disadvantage of these asynchronous time-stepping methods, which affects their numerical stability and accuracy, is their inability to correct predictions for a global timestep without changing the whole hierarchical set of local timesteps.

In addition, these methods have difficulty properly “deactivating” parts of the simulation domain where no physical changes occur. This leads to two undesirable consequences. First, predicting a local timestep is not equivalent to predicting the magnitude of a local *change*. In fact, large changes may occur even for CFL-limited timesteps. This may cause numerical instabilities or lead to poor convergence rates in nonlinear systems. Second, updating variables with hierarchical timesteps may violate *causal relations* between physically related and spatially distributed variables. For instance, in AMR, time stepping typically relies on assigning longer timesteps to larger cells. The larger cells are integrated in time *prior to* the smaller cells. The smaller cells then receive their temporal boundary information from the larger cells. Therefore, updates of *the slower evolving processes at the larger scales* precede updates of *the faster processes at the smaller scales*. As a result, this (block-structured) time stepping may lead to numerical instabilities when used for integration of strongly nonlinear, rapidly evolving multiscale systems.

13.1.3 Asynchronous Time Integration Using Discrete-Event Simulation

An alternative concept to time stepping is Discrete-Event Simulation (DES), or in other words, *event-driven simulation*. DES is particularly well adopted in computer science (e.g., [7, 8]) where this methodology has been used to model temporal evolution of networks and engineering systems whose elements are connected to each other through deterministic or stochastic relations. In these systems events refer to local (quantitative or qualitative) changes to the global state of the system. In general, events may be asynchronously executed (processed) in their scheduled order, or preempted based on conditional statements.

DES was used to enable asynchronous simulations of relatively simple continuous and discrete physical systems by associating events with updates that result in small, physically meaningful changes [9–11]. However, adapting DES to complex physical systems, such as described by systems of partial differential equations (PDEs) or line-discretized equations (e.g., biological networks), is not immediately obvious or guaranteed. This is because executing asynchronous updates solely based on their predicted time increments provides no guarantee that the resulting solution will be stable, meet the required accuracy requirements, and/or even preserve the underlying conservation laws [4].

EMAPS is a *self-adaptive*, “intelligent” version of DES, which resolves these numerical issues and provides an effective approach to parallelization of event-driven simulations based on physical principles. In what follows, we describe the philosophy and algorithmic structure of EMAPS, provide examples, and trace the evolution of this technology from the early one-dimensional “proof of principle” demonstrations to the recent multi-dimensional hybrid simulations of large-scale magnetized plasmas.

13.2 EMAPS: Event-driven Multi-Agent Planning System

13.2.1 Basic Algorithm

To incorporate intelligence into DES, EMAPS adds another layer of complexity by introducing (1) self-adaptive preemption (correction) of events for accuracy and stability, and (2) forced preemption of events with close timestamps for efficient parallelization. In essence, instead of forcing variables in a numerical model to step in time synchronously, EMAPS evolves them by satisfying causal and accuracy constraints derived directly from physical considerations. This *change-based* approach makes EMAPS drastically different from the above-mentioned asynchronous time-stepping algorithms. DES was first adapted and applied to spherically symmetric electrostatic plasma simulations [12]. EMAPS was introduced [4, 13] as a general algorithm for modeling multiscale evolution systems. The acronym itself (“EMAPS”), however, was coined later to emphasize two new features of this approach compared to DES: agent-based intelligence and preemptive event processing (PEP), introduced in [14].

The general philosophy of EMAPS may be explained as follows. First and foremost, instead of advancing the global state of the system by choosing predetermined time increments, EMAPS evolves the system in simulation time by tracking finite changes to its local states (variables) through *the use of local agents*. These agents, by predicting (scheduling) and interrupting (preempting) updates of system states, ensure that all changes caused by these updates (events) always remain physically meaningful. In particular, (1) local states are updated “if and only if” their changes exceed some predefined or adaptively computed thresholds, and (2) each update results in a *small physical change*. The former makes EMAPS computationally efficient, while the latter increases numerical accuracy and convergence when modeling strongly nonlinear and stochastic systems. Conservation laws in EMAPS are preserved by delaying flux updates between cells with common interfaces, as explained below. Particles in EMAPS-driven applications are handled asynchronously, being pushed on their own time scales in self-consistent force fields which may also be updated asynchronously by EMAPS (see Sect. 13.4).

There exists a great and rather obvious difference between how computational efficiency is achieved in the event-driven and time-stepped codes. Compared to “blind” time stepping, for which code optimization is often straightforward, the intelligent calculations performed by EMAPS inevitably incur additional CPU overhead (roughly ~100%), in a way similar to overhead due to extra operations on unstructured meshes (vs. regular meshes). Traditional performance metrics, e.g., as the number of numerical operations per unit time, however, often become irrelevant when comparing event-driven and time-stepped simulations. In realistic multiscale simulations, where computational quantities may dynamically vary by orders of magnitude, the extra overhead due to EMAPS becomes insignificant compared to its main advantage. For instance, for a given simulation time interval, an optimized time-stepped code may have to update, say, 1 million particles at *all cells*, but during the same time EMAPS may need to update only 1000 particles at

some cells, while leaving the particles at other cells “idle” (not updated) because EMAPS finds that these particles do not even need to be updated for a given accuracy. Moreover, under certain physical conditions time-stepped simulations of global, strongly coupled systems (e.g., the Earth’s magnetosphere) may become computationally prohibitive, or even not feasible because of numerical instabilities that often arise due to the general inability of explicit time-stepping models to timely adjust global timesteps in accordance with their nonlinear dynamics. Choosing a proper timestep in such models becomes not trivial due to a generally difficult task of optimizing accuracy vs performance. EMAPS automates this task by letting a user of the event-driven model decide only on *physical tolerances* (maximum allowed changes to computational variables), while EMAPS adaptively determines local time increments in accordance with the user-specified precision.

In general, EMAPS sets forth the following objectives: (1) provide efficient and stable processing of computational variables, (2) eliminate unnecessary computation, (3) increase accuracy and convergence by integrating numerical variables with *self-adaptive update rates* based on local physical and mesh properties; (4) offer a universal framework for including *different physical approximations* within a single simulation model. These objectives are met through predicting, preempting, and synchronizing computational updates in spatial domains of arbitrary topology. To explain this methodology for physics-based simulations, we introduce the following terminology (also see Fig. 13.1):

1. A discrete variable, f represents a *local state* defined at a computational cell. The global system state at any simulation (*global clock*) time is represented by an ensemble of local states defined at their own times (referred to as *timestamps*) which may be different from the global clock time. An update of a local state by its *agent* (local algorithm) is called a *local event*. Global events represent synchronous updates of the global state (e.g., I/O, etc.).
2. Events are *scheduled*, i.e., logically arranged for their execution (*processing*) in an appropriate priority queue (*EventQueue*) in accordance with their timestamps, t_e , so that events with smaller timestamps are processed earlier. Event *preemption* is a mechanism for withdrawing a currently scheduled event before its scheduled execution time. The *engine* is a global scheduler agent that advances the computational model in simulation time by scheduling and executing local and global events.
3. Depending on actions of their agents, local states are classified as *updated* or *synchronized*. A state is considered to be *synchronized* when following its update, its agent forces the engine to assign a new execution timestamp to it (i.e., *reschedule* it) in the event queue. Otherwise, this state is considered to be *updated*. Agents, ordered by the engine to update their states when the engine pops the top event out of the event queue, make their states *directly synchronized*. A directly synchronized state interacts with dependent states by executing *synchronization protocols* via their agents. An agent contacted during a synchronization protocol first updates its state and change detector (*flux capacitor*) and then decides whether to invalidate (*preempt*) its *pending*

- event in *EventQueue* (Fig. 13.1). If the agent decides to do it, it marks its state as synchronized. Preempted (i.e., preemptively synchronized) states continue broadcasting synchronization protocols. This mechanism is similar in spirit to the “corrector” concept widely used in time stepping. It makes EMAPS *self-adaptive* because the communicating local agents satisfy causal relations, through which *local changes to the global system are accurately recorded*.
4. The current simulation time, t_{clock} is called the *global clock time*. The global clock in the system is advanced by the engine in irregular intervals upon processing (in general) a group of events at the current simulation time.
 5. A time increment, Δt_e for a state, f is based on its *target threshold*, Δf , which limits the magnitude of a physical change of state f between its successive processing events. Δf may be a constant tolerance, or a function of f . For accuracy, Δf should not exceed the CFL “quantum” value, $\Delta f_{\text{CFL}} = |\partial f / \partial t| \Delta t_{\text{CFL}}$, where Δt_{CFL} is the CFL timestep for this state. Δf may also be subject to other physical constraints, e.g., those that may not allow changes with magnitudes above specified thresholds (needed, for instance, to accurately model specific reactive terms or particle gyro-motion). Interestingly, it was discovered that for the diffusion-advection-reaction equation, Δf can even be chosen to exceed Δf_{CFL} by orders of magnitude without incurring a numerical instability [4]. Choosing $\Delta f > \Delta f_{\text{CFL}}$, however, leads to noisy oscillations of the order of $\sim \Delta f$ in f . Still, by doing so one may obtain *an approximate solution* extremely fast. On the contrary, it is well known that explicit time stepping with $\Delta t > \Delta t_{\text{CFL}}$ is explosively unstable, which precludes such calculations.

Below we present a serial pseudocode for the EMAPS algorithm that solves a generalized equation, $\partial f / \partial t = R(f, x, t)$. This algorithm is further explained in a line-by-line manner. Its extension to parallel execution is discussed in Sect. 13.2.2.

```
function run_EMAPS()
1: Finished = false;  $t_{\text{clock}} = 0$  // initialize global simulation clock
2: for  $f_i$  in DESFab: // loop over grid-based data
3:    $t_i = t_{\text{clock}}$  // state timestamps are initialized
4:    $f_i.\text{initialize}()$  // all states are initialized and synchronized
5:   add  $f_i$  to PEPStack // stack of events
6: endfor
7: if (TimeOrder > 1 and  $t_{\text{clock}} > 0$ ): // if second order of temporal accuracy chosen
8:   for  $e$  in PEPStack:  $e.\text{correct}()$ 
9: endif
10: if (Finished is true): return // after correcting solution
11: for  $e$  in PEPStack:
12:    $e.\text{dfdt}()$  // compute rate-of-change
13:   if  $e.\text{isSynchronized}()$ :  $e.\text{schedule}()$  // add new event to EventQueue
14: endfor
15:  $t_{\text{top}} = \min(t_e, \infty)$  // minimum valid timestamp in EventQueue
```

```

16:  $t_{clock} = \min(t_{END}, t_{top})$  // set the clock to the earliest time
17: if ( $t_{clock} == t_{END}$ ): Finished = true
18: clear PEPStack
19: while EventQueue is not empty: // do PEP cycle
20:    $e = EventQueue.top()$  // get the earliest valid event
21:   if ( $t_e > t_{clock} + \Delta t_{PEP}$  and Finished is false): break // leave PEP cycle
22:    $e.process()$  // self-adaptively update and synchronize states, compute  $\Delta t_{PEP}$ 
23:    $t_e = t_{clock}$  // update state timestamps
24:   add  $e$  to PEPStack
25:   EventQueue.pop() // discard event  $e$ 
26: endwhile
27: goto line 7 // continue simulation
endfunction

```

Lines 1–6: Variable initialization All cell states, f_i are stored in a data container, *DESFab*. Upon their initialization, they are added to a linked list (*register*), *PEPStack*. At this point all states are considered to be synchronized by the engine, i.e., eligible for scheduling.

Lines 7–9: Second-order flux correction (not done following the initialization stage) The first-order Euler time integration scheme corresponds to *TimeOrder*=1. The second-order asynchronous flux-conserving Euler correction was introduced first for gas dynamics equations [14]. To implement this correction, one needs to define separate cell-centered timestamps for source terms and additional (face-centered or edge-centered) timestamps for flux updates at cell interfaces. The correction algorithm proceeds by replacing the “old” (predicted) change to f by the “new” (corrected) one. This procedure is carried out for all states in *PEPStack* (t_1 and t_2 are the old and new timestamps, respectively):

$$f_{\text{old}}(t_2) - f(t_1) = \int_{t_1}^{t_2} R(t_1) dt \rightarrow f_{\text{new}}(t_2) - f(t_1) = \frac{1}{2} \int_{t_1}^{t_2} (R(t_1) + R(t_2)) dt.$$

Line 10 End the simulation if it has reached the finish time.

Lines 11–18: Event prediction and scheduling All states in *PEPStack* recompute their rates of change, $\partial f / \partial t \equiv R$. The rate of change, R is used for evaluating conditional statements (*change detectors*) in synchronization protocols, as explained below. For synchronized states, this value is also used to predict local time increments, Δt_e and execution timestamps, $t_{clock} + \Delta t_e$ for the corresponding events in *EventQueue* (“predictor”). To reduce the number of *wake-up* events (i.e., events preempted due to detected large changes), the minimum value of Δt_e , computed

in the neighborhood of a given state (cell), is used. Different local strategies make EMAPS flexible and extendable to specific applications. Upon scheduling events, the global clock time, t_{clock} is set to the earliest of the top timestamp in *EventQueue* and the finish time. At this point, *PEPStack* is made empty (line 18). Note that the engine does not predict events with processing times that exceed the finish time. These states are *deactivated*. Although cells can be deactivated in time-stepped simulations as well, time stepping does not offer a general algorithm *for their proper reactivation*. EMAPS makes it straightforward: states are automatically reactivated once they are synchronized and their new scheduling conditions are met. This approach avoids numerical instabilities and leads to significant speed-ups in simulations where the number of active states remains relatively small compared to their total number.

Lines 19–26: Preemptive Event Processing (PEP) Cycle The PEP cycle [14] is performed by the “central processing unit” of EMAPS, the engine. It processes and pops the earliest events out of *EventQueue*. If *EventQueue* is found to be empty, the engine finishes the simulation. Otherwise, after local agents return control to the engine (see the description of the synchronization protocol above), the engine continues to process new events within a *PEP time window*, Δt_{PEP} . This concept measures event “closeness in time.” PEP enables efficient parallel processing of events with close timestamps [14]. Particular implementations of Δt_{PEP} may vary. Currently, Δt_{PEP} is dynamically set to within a constant factor from the minimum difference between successive execution times of states being synchronized by PEP. The PEP cycle is considered to be *completed* when the top event’s scheduling time exceeds the *floating* time, $t_{clock} + \Delta t_{PEP}$. Therefore, at each time, t_{clock} PEP selects and processes a batch of events in general. This makes parallel EMAPS straightforward compared to conservative and optimistic approaches used to parallelize traditional discrete-event simulations (e.g., see [8, 15]).

All states, being processed in a given PEP cycle (line 22), set their timestamps to the current clock time (line 23). A local state, processed directly by the engine, broadcasts synchronization protocols to its dependent states. Upon being contacted by this protocol at $t = t_{clock}$, each dependent agent does the following: (1) updates its state, $f = f + \delta f$, $\delta f = R(t_e)(t_{clock} - t_e)$ using appropriate timestamps for different geometric elements (faces, edges, volumes), and (2) updates its *flux capacitor* (change detector), $\delta f_{flux} = \delta f_{flux} + \delta f$. δf_{flux} is reset to zero when a new event for this state is scheduled. If the absolute value of δf_{flux} in the state being probed exceeds its predicted threshold value, Δf , then this state is considered to be synchronized and a new synchronization protocol is recurrently called. All synchronized states, directly processed by the engine or preempted by local agents through change detection, are placed in *PEPStack*. These synchronization protocols incur some performance overhead. We have found, however, that in multiscale applications the average number of states synchronized per PEP cycle remains relatively small compared to the total number of states (variables) in the system. This leads to

significant performance gains. Importantly, by building self-adaptive trajectories for local variables based on their predicted changes and a posteriori corrections, EMAPS evolves the global model in simulation time in a manner that prevents numerical instabilities that may arise in explicit time-stepped simulations under similar physical conditions [16].

13.2.2 Parallelization

Efficient parallelization of event-driven simulations is not straightforward. Traditional discrete-event applications use various deterministic and optimistic synchronization schemes [8]. Deterministic algorithms always execute events in the event timestamp increasing order, which typically results in too few operations per global synchronization step. Optimistic strategies allow processors to execute their local events ahead of the global clock time. These methods, however, suffer significant performance penalties in situations where rollbacks of parts or the whole of the global system need to be performed [17].

For physics-based simulations, where events represent updates based on physical equations, we resolve these issues by applying the Preemptive Event Processing (PEP) procedure [14], described above (note, however, that not all types of discrete events can be preempted in general DES). PEP enables efficient parallelization of EMAPS-driven models by *synchronously* executing events with close timestamps *prior to their scheduled execution times*. When a simulation model contains distinct groups of events with fundamentally different timescales, separate event queues (and the corresponding PEP cycles) may be implemented for efficiency [18]. For instance, this approach is adopted in a hybrid code, HYPERS (Sect. 13.4), where the “dual PEP” algorithm [18] processes field and particle events in separate event queues on their own characteristic timescales.

It should be emphasized that EMAPS effectively generalizes time stepping by *adaptively updating* parts of a global system at *adaptively selected* time moments. When all events evolve on *a single timescale*, EMAPS automatically reduces event processing to uniform time stepping. PEP offers a general way of synchronizing update events with differing fine- and coarse-grained timescales, commonly encountered in scientific and engineering applications. In this capacity, PEP may be also used as a “universal glue” for combining *numerical models* with disparate timescales into a single parallel application.

In a distributed-memory parallel processor environment, the synchronization protocols have to be applied across the processor shared memory boundaries. This happens, for instance, when a border state of one processor synchronizes itself via message-passing with a dependent state residing on another processor. These wake-up calls may trigger additional parallel communication and overhead. This overhead, however, can be avoided by forcing agents at *all border (communication) cells* to check for their possible flux capacitor overflow *before* the inter-processor communication is initiated. As a result, synchronization messages that arrive at their destination border cells following a PEP cycle do not launch synchronization proto-

cols because all possible synchronization sequences have already been precomputed [18].

Similarly to parallel particle simulations, computational efficiency of PEP-driven applications with inhomogeneous event rates and multiple streams (queues) of asynchronous calculations relies on using proper parallel load balancing strategies. For instance, in order to dynamically compute optimum parallel partitions, global computational load maps may be constructed at runtime [18]. Likewise, using “stretched processor grids” significantly boosts parallel efficiency of magnetospheric simulations performed on stretched physical meshes [19]. Parallel load balancing remains an active research topic in high-performance computing and further advances in this area will have direct impact on parallel efficiency of EMAPS-driven applications.

13.3 Early Applications

To establish the viability of EMAPS for scientific computing, we first applied this novel approach in one dimension (1D) and compared its results against equivalent time-stepped simulations [4, 13, 14], as summarized below. In these examples we used uniform meshes in order to disentangle temporal updates from spatial details. However, it should be emphasized that nonuniform meshes offer even more numerical advantages for EMAPS-driven computing [19].

13.3.1 Diffusion-Reaction-Advection Equations

In general, a one-dimensional flux-conservative (scalar) equation with a source term can be written in the following form:

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial x} \Phi(f, x) = S(f, x). \quad (13.1)$$

Here $f(x)$ is the solution of interest on a spatial domain, $x \in [0, L]$, and $\Phi(f, x)$, $S(f, x)$ are the flux and source functions, respectively. In what follows we assume the presence of diffusion and advection terms:

$$\Phi(f, x) = \Phi_d + \Phi_a, \Phi_d = -D(f, x) \frac{\partial f}{\partial x}, \Phi_a = u f(x), \quad (13.2)$$

where $u > 0$ is a constant advection velocity and $D(f, x)$ is a variable diffusion coefficient. In the discrete model considered, the diffusion term was approximated with central differences and the advection term was discretized with a first-order upwind scheme. The EMAPS integration was performed using a first-order Euler method [4].

This first application of EMAPS immediately led to an interesting observation: self-adaptive integration remains numerically stable and converges to steady-state solutions even for large change thresholds, $\Delta f \gg \Delta t_{\text{CFL}}$ [4]. This result emphasizes the resilience of EMAPS to numerical instabilities that commonly plague time-stepping methods. Indeed, any explicit time stepping is well known to explosively terminate when $\Delta t > \Delta t_{\text{CFL}}$. On the contrary, EMAPS-driven simulations, even when physical changes *greatly exceed their local CFL values* ($\Delta f \gg \Delta t_{\text{CFL}}$), still produce reasonable numerical solutions, albeit with noisy features of the order of Δf . Therefore, even though for time-accurate results one still needs to select $\Delta f < \Delta t_{\text{CFL}}$, it is still possible to get a noisy solution *very fast* with $\Delta f > \Delta t_{\text{CFL}}$ *without suffering catastrophic consequences*. Importantly, time-accurate ($\Delta f < \Delta t_{\text{CFL}}$) solutions of diffusion-reaction-advection equations with variable and nonlinear coefficients were obtained with EMAPS at significant CPU time savings compared to the corresponding CFL-limited time-stepped calculations [4].

For instance, an event-driven solution of Fisher's equation ($D = 0.01$, $S = \gamma f(1-f^2)$, $\gamma = 100$, $u = 0$) was obtained in [4]. This diffusion-reaction equation describes a general class of "pulled front" problems, which typically occur in systems where steep wave fronts enter regions occupied by unstable physical equilibrium states. This particular equation has a smooth analytic solution in the form of a propagating "heat wave," which maintains an almost constant profile in the upstream region. Equations like this one are commonly solved with implicit techniques using large timesteps. In the EMAPS-driven computation, however, the neighborhood of the moving front is the only region where numerical updates are generated. The inactive parts of the computational domain (i.e., cells with negligible update rates) are automatically deactivated [4]. Therefore, EMAPS may also be used as a front-tracking technique. This feature makes it well suited for computing moving shocks and other discontinuities (e.g., flame fronts, radiation diffusion shocks, etc.).

13.3.2 Computational Gas Dynamics

The second-order self-adaptive integration was first demonstrated by solving Euler's equations [14] that form the basis of many computational gas dynamics models:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{U})}{\partial x} = 0, \quad (13.3)$$

$$\mathbf{U} = \begin{bmatrix} \rho \\ M \\ E \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} M \\ \rho v^2 + p \\ (E + p) v \end{bmatrix}, \quad E = \frac{p}{(\gamma - 1)} + \frac{\rho v^2}{2}. \quad (13.4)$$

In Eqs. (13.3) and (13.4) \mathbf{U} and \mathbf{f} represent the solution and flux vectors; ρ , \mathbf{v} , $M = \rho\mathbf{v}$, p , E are the density, velocity, momentum, pressure, and total energy of gas, respectively; and γ is the gas adiabatic index. These equations were discretized in space using a second-order upwind-central scheme and evolved in time with local $\Delta f \leq \Delta t_{CFL}$. The second-order flux correction produced solutions that matched reference time-stepped solutions with second-order temporal accuracy. An example of automatic activation/deactivation of cells was demonstrated in the simplest case of constant-velocity advection. In addition, the ability of EMAPS to adjust update rates in accordance with local CFL rates in complex gas dynamics problems was also demonstrated [14].

In all cases, the event-driven solutions were found to closely match the corresponding solutions obtained with time-stepping schemes. The asynchronous solutions, however, were obtained at significant CPU time savings, proportional to the amount of eliminated computation. In general, this speed-up due to EMAPS increases with a larger problem dimensionality and degree of system inhomogeneity.

13.3.3 Hybrid Simulations of Fast Plasma Shocks

One of the most useful plasma approximations is a quasi-neutral hybrid model (see [20] and also chapters in this book for different hybrid model implementations). The traditional hybrid model treats the plasma ions as kinetic species using Particle-in-Cell (PIC) or Vlasov [21] descriptions, neglects displacement current in Maxwell's equations, assumes plasma quasi-neutrality, and approximates the plasma electrons as a massless adiabatic fluid (optionally, an electron pressure equation may be solved as well):

$$\mathbf{E} = \frac{\mathbf{j}_e \times (\mathbf{B} + \mathbf{B}_{ext})}{en_e c} - \frac{\nabla p_e}{en_e} + \eta \mathbf{j}, \quad (13.5)$$

$$p_e \sim n_e^\gamma, en_e = \rho_i, \quad (13.6)$$

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{j}, \mathbf{j} = \mathbf{j}_e + \mathbf{j}_i, \quad (13.7)$$

$$\frac{\partial \mathbf{B}}{\partial t} = -c \nabla \times \mathbf{E}. \quad (13.8)$$

In Eqs. (13.5)–(13.8) n_e , \mathbf{j}_e , p_e are the electron number density, current density, and pressure. Further, c is the light speed; e is the absolute value of the electron charge; \mathbf{E} , \mathbf{B} are the self-generated electric and magnetic fields; \mathbf{B}_{ext} is the external magnetic field; ρ_i , \mathbf{j}_i are the ion charge density and current densities; η is the plasma resistivity. Equations (13.5)–(13.8), together with Newton-Lorentz's equations of motion for ion macro-particles, are known to describe a broad range of plasma

waves with potentially disparate timescales. Among the fastest and most restrictive scales are those generated by short-wavelength whistler oscillations and energetic particles. Time-stepped hybrid codes choose global timestep values based on the most restrictive CFL conditions for particles and fields. However, even elaborate time-stepping schemes for Eqs. (13.5)–(13.8) and equations of ion motion may produce explosive numerical instabilities in situations where the combination of nonlinear effects and particle noise invalidates the viability of linear stability conditions used for estimating global timesteps.

Following the demonstration of event-driven algorithms for particle systems and partial differential equations (Sect. 13.3.1), a first-ever self-adaptive hybrid code was constructed in one dimension [13]. This hybrid code used the Nearest-Grid-Point (NGP) interpolation scheme, previously employed in an event-driven 1D electrostatic code [12]. In contrast to the electrostatic simulations, however, the electromagnetic hybrid codes perform time-accurate integration of electromagnetic fields. To numerically preserve the divergence-free property of magnetic field ($\nabla \cdot \mathbf{B} = 0$) in this 1D hybrid model, the magnetic field, \mathbf{B} was represented by the vector-potential, $\mathbf{A}(\mathbf{B} = \nabla \times \mathbf{A})$, which was advanced in time with a first-order event-driven scheme. An ad-hoc plasma resistivity was also applied to suppress excessive particle noise produced by the NGP scheme. This event-driven hybrid code was used to simulate fast plasma shocks and concomitant ion acceleration. The event-driven simulations were also compared in detail with the corresponding explicit and implicit conventional simulations [13].

In addition to achieving the expected performance gains, these first event-driven hybrid simulations produced a number of interesting results. First, the NGP-based event-driven algorithm was found to be numerically stable, while two other time-stepped hybrid codes required linear interpolation to maintain their numerical stability. Second, not only the event-driven simulation was found to closely match the temporal dynamics and spatial profiles of low-Mach-number and intermediate plasma shocks in the corresponding time-stepped simulations, but EMAPS also produced better resolved profiles of high-Mach-number fast magnetosonic shocks. Four zones of shock-driven turbulence were clearly resolved in the event-driven simulation [13]: (1) low-frequency steepened oscillations observed far upstream (“shocklets”), (2) short-wavelength oscillations in the near upstream region (driven by the reflecting ions), (3) the coherent shock transition region where the upstream oscillations get compressed and amplified, and (4) long-wavelength (rarefaction) waves in the downstream region. In addition, the applied resistive smoothing enabled EMAPS to generate larger timesteps, which contributed to overall performance. A similar effect was observed in more recent 3D HYPERS simulations of field-reversed configurations with ion–ion collisions [22].

It was later found that event-driven hybrid simulations also produce more accurate and fast converging particle energy spectra compared to similar time-stepped simulations of shock-driven ion acceleration [9]. This illustrates the ability of an EMAPS-driven hybrid simulation to accurately follow nonlinear plasma dynamics and ion kinetics. It should be also noted that applying unnecessarily small timesteps in a time-stepped hybrid simulation inevitably increases numerical

diffusion. On the other hand, simulations with larger global timesteps may miss important local particle–field interactions, automatically followed by EMAPS [9].

13.4 HYPERs: Hybrid Parallel Event Resolving Simulator

HYPERs is an EMAPS-driven, compile-time configurable (2D/3D in configuration space), massively parallel hybrid code [16]. Particles and electromagnetic fields in HYPERs are asynchronously updated on logically uniform meshes which may be uniform or stretched in physical space [19]. In addition to asynchronous local updates (local events), HYPERs also performs synchronous global operations (global events), such as input/output operations, diagnostics, particle collisions, and particle injection events, among others. Field and particle events may be scheduled/processed in a single or separate event queues, depending on the problem.

The magnetic field components, \mathbf{B} are defined on staggered meshes to preserve $\nabla \cdot \mathbf{B} = 0$ with round-off precision. This numerical divergence-free property of magnetic field is an important condition for minimizing numerical diffusion and dispersion errors in hybrid simulations. HYPERs integrates the face-centered magnetic field components, \mathbf{B} using Faraday’s law (13.8). The electric field, \mathbf{E} is found from the generalized Ohm’s law, $\mathbf{E} = \mathbf{E}(\mathbf{B}, p_e, n_e, \mathbf{j}_i, \mathbf{j})$, obtained by combining Eqs. (13.5)–(13.7). For a given cell (*F-state*), a “field event” updates its face-centered magnetic components, together with its node-based and edge-based electric field components. Local time increments for field events are predicted by calculating the minimum of two values, $\Delta B = \min(\Delta B_{\text{CFL}}, \Delta B_{\max})$, where ΔB_{\max} is a user-defined value, and $\Delta B_{\text{CFL}} = \Delta t_{\text{CFL}} |\partial B / \partial t|_{\max}$, where $|\partial B / \partial t|_{\max}$ is the maximum absolute value computed component-wise, and Δt_{CFL} is the minimum CFL value found from the linear dispersion laws applicable to the hybrid model. Based on the computed ΔB , EMAPS estimates a local time increment, $\Delta t = \Delta B / |\partial B / \partial t|_{\max}$. The final time increment for a given *F-state* is predicted by taking the minimum of all predicted values, Δt computed in the neighborhood of this state’s cell.

Macro-particles are stored in cell-based local priority queues (*P-states*), sorted according to *particle timestamps*. A particle timestep is predicted by (1) restricting an allowed particle displacement to a fraction of its physical cell size (this fraction is currently constant but in general it may be a function of cell position), and (2) imposing local gyrofrequency restrictions. More constraints may be imposed in general, if needed. At each cell, particles of different species are grouped into different *P-states*. Therefore, a particle event at a given cell updates only particles in the appropriate *P-state*, not all particles at this cell. In addition to updating particle positions and velocities, particle events also perform bookkeeping operations that place particles at their destination *P-states*. A particle event is always scheduled for execution at the earliest particle timestamp found in its *P-state*. When processed, it executes its own *mini-PEP* algorithm, similar to the global PEP algorithm used for preempting selected events in event queues (Sect. 13.2). Likewise, this *mini-PEP* algorithm in general updates only a fraction of all *P-state* particles. Synchronized

P-states, which have their predicted execution timestamps changed during particle update operations, are rescheduled upon execution.

Currently, HYPERS uses linear interpolation for scattering a particle charge and current to the nearest mesh nodes, as well as interpolating field values from these mesh nodes to the particle position. Each asynchronous particle push requires two scatter operations: subtraction of old particle moments and addition of new ones. Future code implementations may optimize this algorithmic step by taking advantage of charge-conserving interpolation schemes which compute particle-induced mesh currents in a single push operation.

13.4.1 Magnetic Field Correction

The asynchronous time integration scheme, implemented in the original version of HYPERS [16], identically preserves the numerical analog of $\nabla \cdot \mathbf{B} = 0$ on Yee's mesh in the first-order temporal approximation. The second-order correction used in [16] is, however, generally non-conservative. Below we describe a straightforward algorithmic fix to this issue, adopted in the later versions of HYPERS.

In the original algorithm the second-order magnetic field components, \mathbf{B}_2 at a given cell were obtained from the corresponding first-order accurate solution, \mathbf{B}_1 using a *topologically inconsistent* operation:

$$\mathbf{B}_2 = \mathbf{B}_1 - \Delta t_B \nabla \times \delta \mathbf{E}, \quad (13.9)$$

where $\delta \mathbf{E}$ is the edge-centered electric field correction, $\nabla \times$ is the discrete curl operator for the Yee-cell, and Δt_B represents the *face-centered time increments* for the corresponding magnetic field components. This second-order correction was shown to speed up HYPERS simulations in [16]. However, despite that $\nabla \cdot \nabla \times \mathbf{E} = 0$ numerically holds at any time, the correction expressed by Eq. (13.9) generally violates the discrete analog of $\nabla \cdot \mathbf{B} = 0$ by introducing asynchronous errors of order $\sim O(\Delta B)$. Although these errors appear to be less significant at cells with stronger magnetic fields, $|\mathbf{B}| \gg \Delta B$, they may still produce significant noise at cells where $|\mathbf{B}| \sim \Delta B$ [18]. Since EMAPS treats this noise as a physical phenomenon, it automatically reduces local time increments for its accurate processing, which may also slow down the hybrid simulation.

These asynchronous errors can easily be eliminated by defining *an additional set of timestamps* for the edge-centered \mathbf{E} components and implementing a *topologically conforming* correction [18]:

$$\mathbf{B}_2 = \mathbf{B}_1 - \nabla \times (\Delta t_E \delta \mathbf{E}), \quad (13.10)$$

where Δt_E represents the *edge-centered time increments* corresponding to the edge-centered electric field components, \mathbf{E} . The new correction preserves $\nabla \cdot \mathbf{B} = 0$ within round-off errors and produces numerically well-behaved solutions [18]. These solutions were also found to be less diffusive compared to the corresponding

solutions obtained with the time-stepped hybrid code, H3D that uses a co-located magnetic field layout, which violates the divergence-free property of magnetic field.

13.4.2 Simulation Geometry

HYPERS defines open (resistive), conducting, and periodic boundary conditions for electromagnetic fields and particles, as well as a number of particle initialization and injection schemes for background plasmas and beams. In complex geometries, physical walls are approximated with masked cells that form “staircase” numerical boundaries (see an example in Sect. 13.5.2). More accurate (cut-cell) approximations may be developed in the future, if needed. Note that CFL conditions for smaller (cut) cells typically require smaller time increments. Time-stepped simulations with cut cells have to use smaller timesteps for all cells. In HYPERS, however, this bottleneck may be naturally alleviated by EMAPS.

Until recently, HYPERS simulations had been conducted on simulation domains covered with uniform meshes. A stretched mesh capability has been implemented recently to enable larger computational domains and accurate far-field boundary conditions in global 3D simulations of the Earth’s magnetosphere [19]. By allowing arbitrary box-structured parallel domain decompositions, HYPERS also enables different strategies for creating efficient processor grids where more processors can be focused on more compute-intense parts of the computational domain. Some details of the stretched mesh and domain decomposition algorithms can be found in [19].

A global magnetospheric hybrid simulation code needs to be able to (1) correctly describe the inflow solar wind region with frozen-in interplanetary magnetic field (IMF) and (2) properly handle boundary conditions at the lateral and outflow surfaces of the simulation box. For accuracy, the inflow boundary should be located far enough from the magnetic dipole. In HYPERS, one may add an “image” dipole that cancels the IMF B_x (GSM coordinates) component at the injection boundary, or make the dipole field fall off faster with distance from the Earth without insignificantly modifying the field in the near distance. In combination with a properly stretched mesh, the latter option makes it straightforward to set inflow conditions under which the solar wind accurately carries a prescribed IMF into the simulation domain [19]. In magnetospheric simulations, for particles HYPERS implements semi-reflecting conditions at the lateral boundaries and absorbing conditions at the inflow and outflow boundaries [19]. The semi-reflecting boundaries reflect drifting Maxwellian particles and absorb shock-reflected and energetic particles. High-resistivity layers around all non-periodic boundaries damp outgoing and reflected waves. Periodic conditions are also available in HYPERS for local turbulence studies.

The shorting out of electric field in the Earth’s ionosphere is currently modeled by placing a perfectly conducting sphere around the Earth with a typical radius of several Earth’s radii. This representation is valid for transient phenomena. For longer simulations, the spherical conductor may be assigned a finite resistance in

order to enable flux tube convection across the polar caps. Note that the MHD codes require defining a plasma velocity condition on the inner boundary as a function of the ionospheric electric field. In hybrid simulations, however, particles are directly coupled with self-consistent electric fields in a resistive circuit.

13.4.3 Collisions and Resistivity

HYPERS implements ion–ion energy and momentum conserving collisions needed for simulations of dense laboratory plasmas. These collisions are implemented using a mesh-based Coulomb collision model [23]. Other collisional (e.g., charge exchange) effects may be included in the future as needed by specific applications. In principle, EMAPS allows collisions with strongly variable rates to be modeled asynchronously.

The standard quasi-neutral hybrid model does not include radiation effects and describes only limited electron physics by retaining the Hall and electron pressure terms in the generalized Ohm's law. To enable fast magnetic reconnection and/or fast propagation of magnetic field in vacuum the hybrid model needs to incorporate an ad-hoc (applied) resistivity model. HYPERS implements several resistivity models, used for different purposes. For instance, to model the lunar wake [24], HYPERS defines a smooth high-resistivity profile for the interior of the Moon, a high constant resistivity in the lunar wake (“vacuum”), and a low constant resistivity inside the plasma. The actual resistivity used by the field solver is found then by spatially averaging local resistivity values. For magnetic reconnection studies, HYPERS currently offers two options: a standard power-law (current-based) resistivity model and a Chodura resistivity model, previously used to reproduce magnetic reconnection in laboratory plasma experiments [22]:

$$\eta_{\text{ch}} = \frac{4\pi v_{\text{ch}}}{\omega_{\text{pe}}^2}, \quad v_{\text{ch}} = c_{\text{ch}}\omega_{\text{pi}} \left[1 - \exp\left(-f_{\text{ch}} \frac{v_{\text{d}}}{v_s}\right) \right], \quad (13.11)$$

where v_s is the ion sound speed, $v_d = |\mathbf{j}|/en_e$, ω_{pi} is the proton plasma frequency, and $c_{\text{ch}} \approx 0.1$, $f_{\text{ch}} \approx 0.3$. Using Eq. (13.11), we were also able to reproduce current sheet thinning and reconnection in flow-driven magnetotail studies (unpublished), similar to those demonstrated in full PIC simulations [25].

13.4.4 Programming Structure

The global computational domain in a HYPERS simulation can be decomposed into an arbitrary number of disjoint rectangular boxes (“subdomains”) which can be assigned to parallel processors in an arbitrary manner. Each subdomain is surrounded by a layer of ghost cells. Adjacent subdomains exchange data via local or remote message passing depending on whether source and target subdomains

reside on the same or different processors. The field data at the ghost cells are either updated by applying local boundary conditions or received from the corresponding border states interior to their adjacent subdomains. Similarly, particles are transferred from ghost to border cells or obey proper boundary conditions. The particle and field states selected for PEP updates are organized into separate linked lists, *PPEPList* and *FPEPList*, respectively. The communication of ghost particle states and border field states is facilitated by linking them into additional lists, *SentGhostList* and *SentBorderList* [18]. Outgoing particles (particles migrating to ghost cells) deposit mesh moments upon arrival at their destination (border) cells. Accordingly, before any particle communication takes place, the subdomain interfaces store partial grid moments, which are then summed up with new ghost moments so that the net interface moments can be used in the field solver. To reduce communication overhead, partial particle moments are exchanged by blocks of cells rather than transferred on the cell basis. All message-passing communications are implemented via a unified abstract programming interface (API). Local (shared memory) operations are implemented via simple memory copying, while remote (distributed-memory) data transfers are performed with MPI. Both cell-level and block-level communications are supported. In addition to facilitating parallel programming, this API allows effective debugging of asynchronous runs. A hybrid (MPI/OpenMP) approach may be implemented in the future for more efficient use of parallel processors.

HYPERS utilizes some data structures and functionalities from the Chombo software framework developed at LBNL. These hierarchies of C++ classes abstract geometric, physical, and communication concepts. The object-oriented design of HYPERS allows adding new code features in a systematic and self-consistent manner. The CPU intense computations are programmed using a dimension-independent pseudo language translated to Fortran 77 by Chombo utilities. The memory allocation for field and particle data is dynamic and handled via distributed data structures. The parallel decomposition approaches adopted assume that particles reside on the same processor with cells they occupy. Cache efficiency is boosted by organizing particles into cell-local containers which keep track of empty memory locations so that they can be filled by incoming particles. These particle containers are implemented using templated data structures. HYPERS also provides a memory “garbage collector” routine that can be periodically called to adjust particle containers and free unused memory.

HYPERS periodically saves (via parallel MPI I/O) simulation data for restarts in the form of binary files, with a single file being written for each multi-processor node. For diagnostics and post-processing purposes, simulation data are saved in the form of binary and text (ASCII) data. These data are processed by parallel (MPI-based and OpenMP-based) programs that generate text and binary information for further visualization. For visualization we currently use the commercial IDL language and a publicly available package, Paraview. Simulation input data are read from text files in a convenient symbolic form using a Chombo utility that enables use of heterogeneous data structures (numbers, arrays, strings, etc.).

13.5 Multi-dimensional Simulations of Magnetoplasmas

To visually demonstrate HYPERS capabilities, below we present results from selected simulations of space and laboratory magnetoplasmas. Most of these results are displayed as figures. However, to complement these figures and give the reader a better feel for asynchronous EMAPS-driven simulations, we have also made two sets of computer-generated, time-dependent videos available online (<https://doi.org/10.5281/zenodo.6508304>). The first (“ipshock”) series of movies visualizes a 2D HYPERS simulation of an interplanetary (IP) shock. These movies illustrate the temporal dynamics of the following quantities: field timesteps (“dt”), particle timesteps (“dt0”), adaptive (Chodura) resistivity (“eta”), magnetic field magnitude (“bb”), plasma density (“rho0”), parallel (with respect to the ambient magnetic field) ion temperature (“tl0”), and double perpendicular ion temperature (“tp0”). The “msx1p” series movies visualize a 3D HYPERS simulation of the laboratory (MSX) plasma experiment in a complex geometry [26] (see below). These movies illustrate dynamic field timesteps (“dt”), magnetic field lines (“b”), plasma density (“rho”), and triple total ion temperature (“ti”).

13.5.1 Global Magnetospheric Simulations

Global hybrid simulations of space plasmas (e.g., see [19, 24, 27, 28]) are typically initialized with a uniform solar wind plasma flow streaming past a spherical conducting (“Earth”) or resistive (“Moon”) obstacle. When modeling the Earth’s magnetosphere, this obstacle represents an inner magnetospheric boundary with a magnetic dipole. It may also be surrounded by a cold immobile dense plasma. In addition, polar cap outflows may be enabled to simulate the impact of cold ionospheric ions on the global magnetosphere.

HYPERS implements a variety of solar wind drivers. In particular, interplanetary (IP) shocks with given parameters (Mach number, density, temperature) are launched by changing (1) the velocity distribution function of injected particles—to initialize the shock Mach number and temperature, and (2) particle weights—to initialize the shock density. One can also modify the tangential electric field at the inflow boundary in order to launch a rotational discontinuity (RD). This option rotates IMF components which are tangential to the solar wind injection boundary.

The obstacle may be set to be reflecting, absorbing, or partially reflecting for particles that reach its surface. The size of this obstacle and the magnetopause position typically need to be scaled down in magnetospheric simulations due to computational limitations. For instance, the typical proton inertial length, d_p in the solar wind is of order ~ 100 km, the Earth’s radius is of order $\sim 64 d_p$ and the typical magnetopause distance, R_{MP} is of order $6\text{--}15 R_E \sim 400\text{--}1000 d_p$. Given the typical cell size of the order of d_p , these scales impose severe constraints on the total number of cells needed to cover the whole computational domain. Therefore, the magnetopause radius in current 3D simulations is typically chosen to be $\sim 100 d_p$.

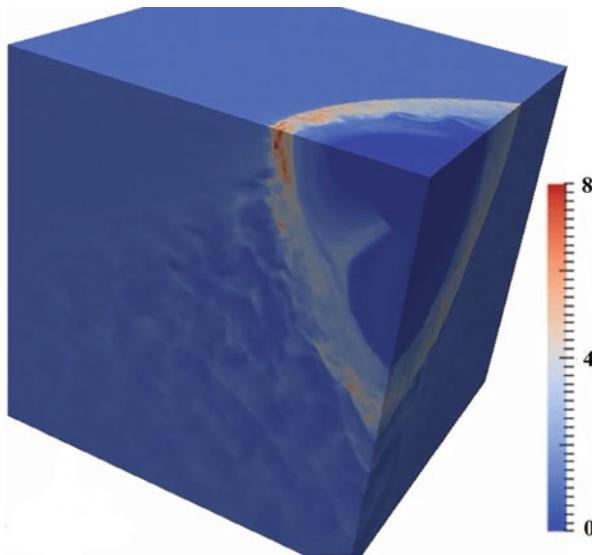


Fig. 13.2 Plasma density in the quasi-parallel shock region in a 3D HYPERS simulation described in [24]. Shown is a quarter of the global simulation domain, $1024 \times 2048 \times 2048 d_p$. The following parameters were used in the simulation: the solar wind Mach number is 10; the IMF angle is $\theta_{\text{IMF}} = 22^\circ$; the proton and electron betas are $\beta_p = 0.6$, $\beta_e = 1.6$, respectively; the obstacle radius is $92 d_p$ and the magnetopause position is $160 d_p$. The simulation was performed on 131,072 processor cores of the NSF's Blue Waters

HYPERS is an efficient computational tool for studying interactions of solar wind with the Earth's magnetosphere and other space bodies, including the Moon. For instance, HYPERS has recently been used to conduct some of the largest 3D global hybrid simulations of the Earth's magnetosphere to date (Fig. 13.2), focused on studying statistical properties of the Earth's foreshock turbulence [24]. The three-dimensional computational domain in this study was composed of $820 \times 1640 \times 1640$ cells ($1024 \times 2048 \times 2048 d_p$).

Global configurations of magnetic and electric fields in the Earth's magnetosphere cause plasma turbulence, magnetic reconnection, and large flux transfer events. These processes drive a broad range of waves that interact with various particle populations. For example, waves can accelerate cold ions of ionospheric origin that can circulate in the global magnetosphere along the interconnected field lines. Building a global multi-dimensional kinetic simulation model capable of predicting the complex multiscale response of the Earth's magnetosphere to solar wind drivers is critical for understanding the physics of foreshock, bow shock, magnetosheath, magnetopause, magnetotail, as well as reconnection effects and response of magnetospheric populations to extreme solar events. The adaptive, multiscale nature of HYPERS is well suited for such simulations, ready to be enhanced with new multi-physics capabilities that will enable even more realistic simulations of the magnetosphere. In this chapter we present results from HYPERS

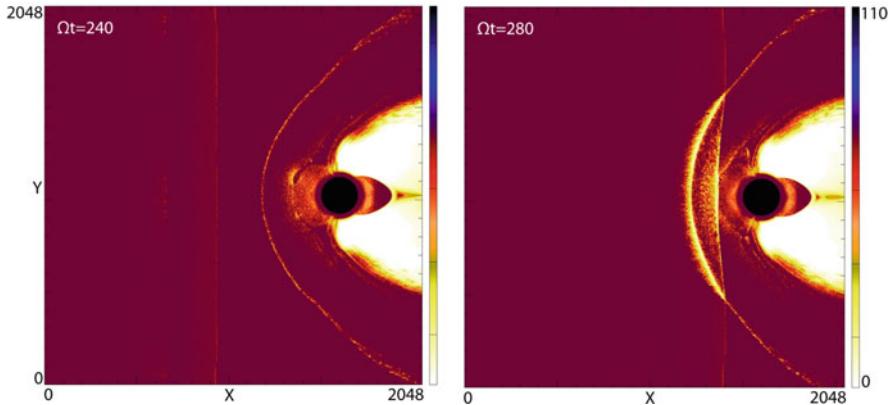


Fig. 13.3 Local field timesteps for two different simulation times (Ω is the IMF-based proton cyclotron frequency) in a 2D HYPERS simulation of the interaction of an interplanetary shock with the Earth’s magnetosphere. The black color indicates inactive cells

simulations performed on uniform meshes only. Global 3D simulations of solar wind interactions with the Moon (lunar wake) and the Earth’s dayside magnetosphere (high-speed jets), including simulations performed on stretched meshes, are discussed in [19, 24].

By conducting extensive head-to-head comparisons with the time-stepped hybrid code, H3D [16], HYPERS has been shown to achieve superior performance metrics in terms of stability, speed, and accuracy (less diffusive and less dispersive). We also note that HYPERS provides stable and converged solutions for cells larger than the ion inertial length. To load-balance parallel computations, HYPERS may use block-structured or stretched “processor grids.” These grids focus more computational resources on the near-Earth region, which reduces execution time by a factor of 3 compared to equivalent uniform processor grids.

Figures 13.3 and 13.4 provide examples for local timesteps from a 2D HYPERS simulation of the interaction of a strong IP shock with the Earth’s dayside magnetosphere. In this simulation, we used the Chodura resistivity given by Eq. (13.11). HYPERS describes this interaction in a truly multiscale fashion by adaptively selecting field and particle timesteps in time in accordance with the magnetosphere’s response to the shock.

Newer space-time capabilities of HYPERS have recently enabled 3D hybrid simulations of high-speed jets (HSJs), frequently observed in the Earth’s magnetosheath [19]. The 3D structure of these jets, as well as the turbulent magnetosheath density variations and magnetic field topologies, is best illustrated in 360-degree videos created for two snapshots with different orientations of the interplanetary magnetic field (<https://doi.org/10.5281/zenodo.4747515>).

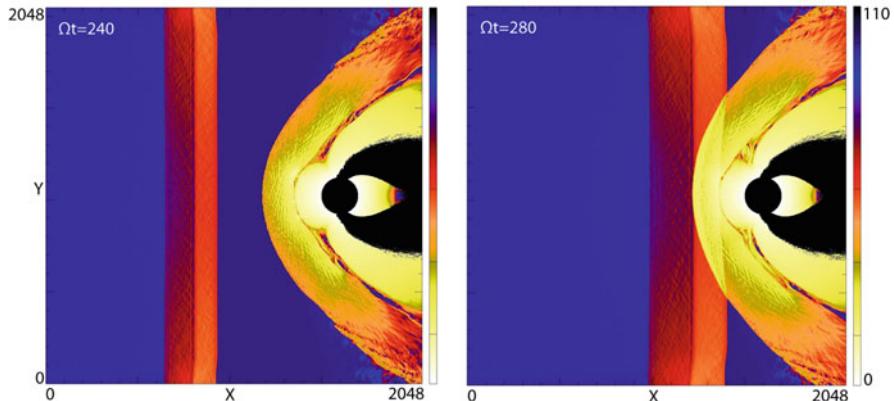


Fig. 13.4 Mesh-averaged particle timesteps for two different simulation times (Ω is the IMF-based proton cyclotron frequency) in a 2D HYPERS simulation of the interaction of an interplanetary shock with the Earth's magnetosphere. The black color indicates inactive cells

13.5.2 Colliding Magnetoplasmas

One of the most challenging goals in plasma physics is to understand physical links between turbulence, coherent plasma structures, reconnection, and dissipation in strongly nonlinear regimes. This can be done in space observations, or in laboratory by studying dynamics of fast evolving plasma “plumes,” initially created in the form of compact spherical magnetized plasma objects (“spheromaks”).

A single spheromak or two double-sided spheromaks can be created at the ends of an experimental cylindrical device and pushed forward at varying speeds. Their dynamics and merging produce small-scale cells of turbulence as well as generate shock structures. These large-scale plasmas are extremely difficult to model with time-stepped hybrid codes because of inhomogeneous and strongly dynamic plasma density and magnetic field scales. To avoid explosive numerical instabilities in simulations, which may arise due to uncontrolled growth of high-frequency oscillations, these dynamic physical timescales must be accurately handled. As an example of how HYPERS overcomes these difficulties, we present results (Figs. 13.5, 13.6, and 13.7) from 3D simulations of colliding spheromaks. These simulations confirm that through merging and reconnection of two spheromaks with opposite helicities one may achieve a stable field-reversed configuration (FRC) [22, 29], provided that the merging spheromaks can arrive at their collision point fast enough, before they may decay into helical structures due to their inherently unstable states.

These simulations were carried out on a four dual-core processor workstation. To achieve this temporal accuracy, however, an equivalent time-stepped hybrid simulation may require hundreds of CPUs, if it can run stably under these stringent conditions at all [30]. The inherent difficulty of these simulations is due to the fact that the numerical timescales, governing the magnetic reconnection and

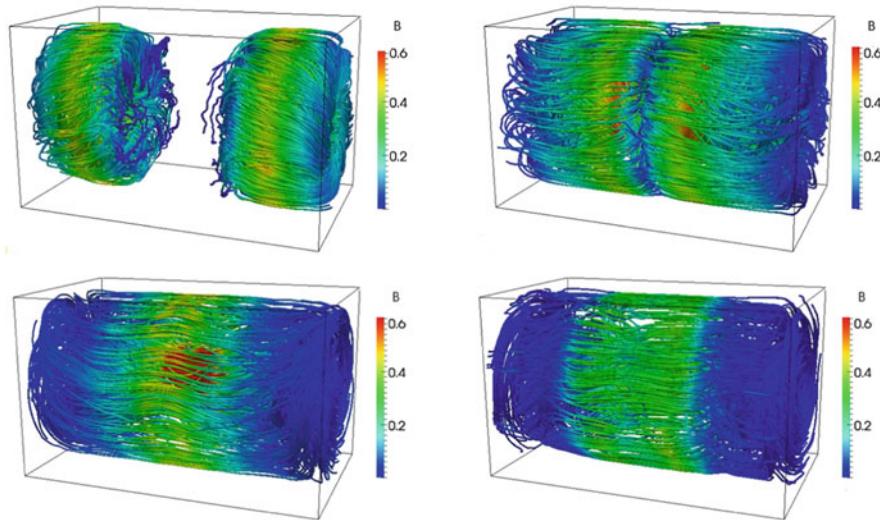


Fig. 13.5 Colliding spheromaks: magnetic field lines. Simulation time progresses from left to right, top to bottom: $\Omega t = 10, 30, 50, 100$, where Ω is the characteristic ion gyrofrequency

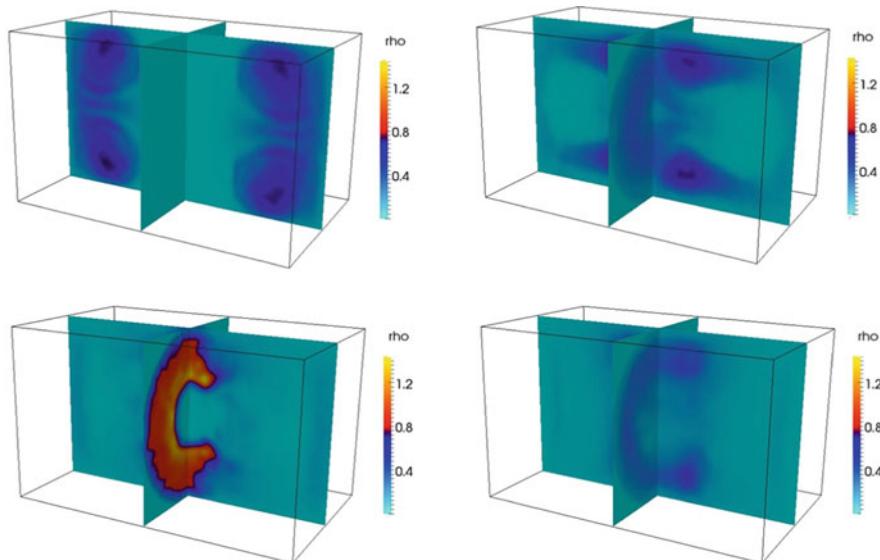


Fig. 13.6 Colliding spheromaks: plasma density. Simulation time progresses from left to right, top to bottom: $\Omega t = 10, 30, 50, 100$, where Ω is the characteristic ion gyrofrequency

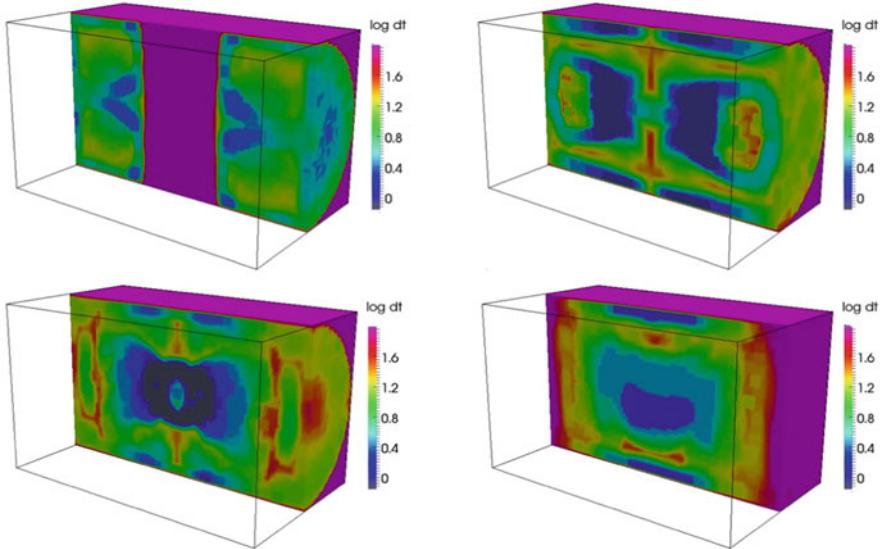


Fig. 13.7 Colliding spheromaks: self-adaptive field timesteps (logarithmic scale). Simulation time progresses from left to right, top to bottom: $\Omega t = 10, 30, 50, 100$, where Ω is the characteristic ion gyrofrequency

concomitant plasma heating in these simulations, vary dynamically in time by orders of magnitude (Fig. 13.7).

3D HYPERS simulations helped interpret experimental results from the MSX experiment at LANL [26] on magnetized collisionless shocks created through the acceleration and subsequent stagnation of FRC plasmoids against strong magnetic mirrors and flux-conserving boundaries (see Figs. 13.8 and 13.9). Local time increments required for field updates in these simulations vary dynamically by two orders of magnitude.

13.5.3 Plasma Expansion Across a Transverse Magnetic Field

Understanding the behavior of plasma flows across strong magnetic fields has direct relevance to many space and astrophysical phenomena (e.g., the Earth's magnetosheath and magnetopause, young stellar objects and astrophysical jets, etc.). Below we demonstrate how published results from two astrophysics-in-lab experiments on laser-produced plasmas helped set up 3D HYPERS simulations that explain the physics of these experiments.

Using parameters of the LAPD experiment [31], HYPERS simulations were set up to model the expansion of a carbon plasma into vacuum across a strong applied magnetic field [32]. These simulations reproduced the temporal dynamics of flute-like density striations that develop at the leading edge of the expanding plasma

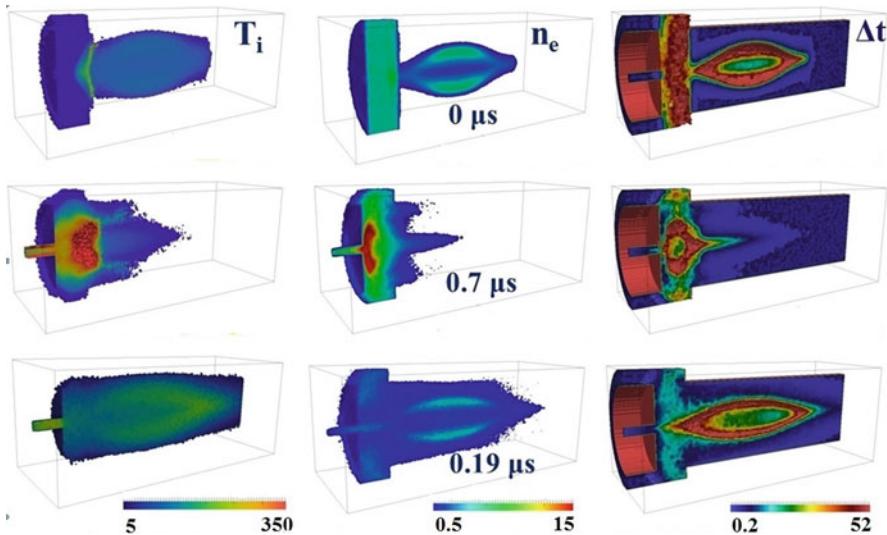


Fig. 13.8 Evolution of ion temperature (left), plasma density (center), and local field timesteps (right) in a HYPERS simulation of the collision of a field-reversed-configuration with a strong magnetic mirror immersed in a pre-ionized plasma in the MRX experiment

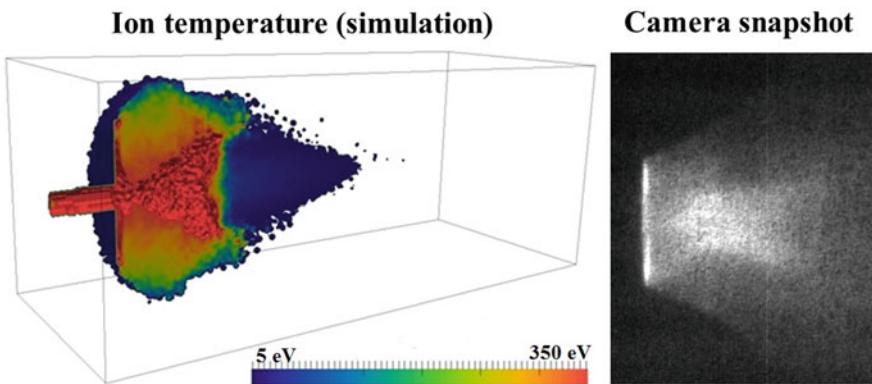


Fig. 13.9 HYPERS simulations reproduced key features of the MRX experiment. Left: an ion temperature snapshot from a simulation. Right: an experimental plasma image

plume due to a large-Larmor-radius Rayleigh-Taylor instability (Fig. 13.10). These simulations also predicted the early destruction of a magnetic bubble in agreement with the experiment, confirming the anomalously rapid decay of magnetic field on timescales ~ 500 ns that are much faster than the magnetic diffusion time ~ 100 μ s.

Another experiment at the University of Nevada, Reno revealed the surprising convergence of an explosively expanding plasma flow into a transversely narrow, collimated stream in the presence of a strong applied magnetic field [33]. Previous

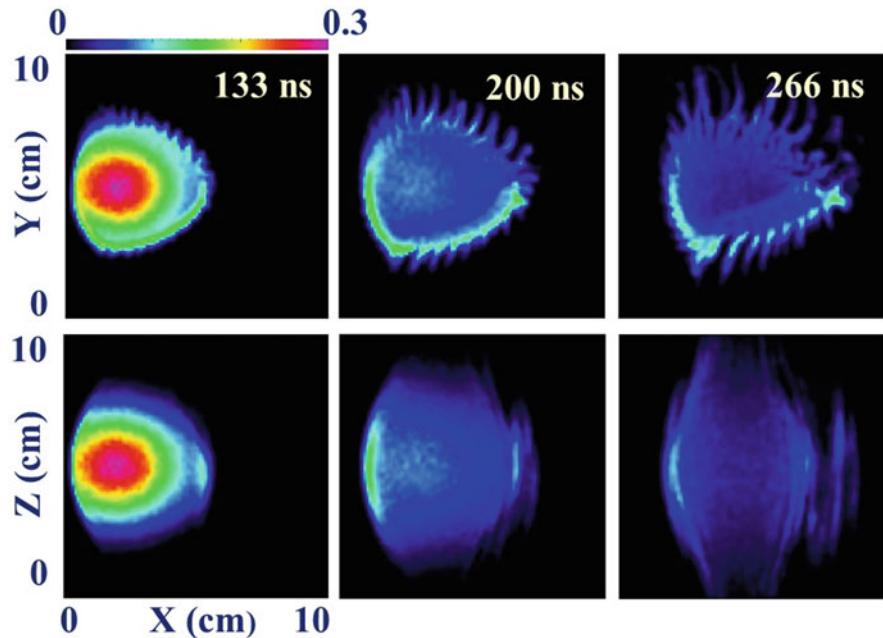


Fig. 13.10 Evolution of plasma density in a HYPERS simulation of the LAPD experiment. A laser-produced plasma expands across a strong magnetic field applied in the Z-direction. This hybrid simulation reproduced flute modes [31], which rapidly develop at the plasma surface and cause early destruction of a diamagnetic cavity (“magnetic bubble”), as observed in the experiment

MHD simulations, which did not take into account the Hall, ion cyclotron and multi-species effects, were only able to predict the lateral plasma confinement but failed to explain the collimated flow. The HYPERS simulation of this experiment, however, exposed the physics of this new phenomenon in great detail. In this experiment, the laser-ablated plasma consisted of protons (H^+) and carbon ions (C^{+4}), since the polyethylene target was initially composed of CH_2 chains. The HYPERS simulation revealed that the energetic plasma, which penetrates into the applied magnetic field, is entirely made of protons. At the same time, the heavier carbon ions lag behind the fast collimated proton stream (Fig. 13.11), composing the bulk of the rear plasma. This ion species separation effect could not be deduced from experimental photographs which recorded only the electron density. The HYPERS simulation also indicates that the main mechanism for this fast plasma penetration is a combination of plasma $E \times B$ drift and magnetic diffusion, where the transverse electric field arises due to the Hall effect.

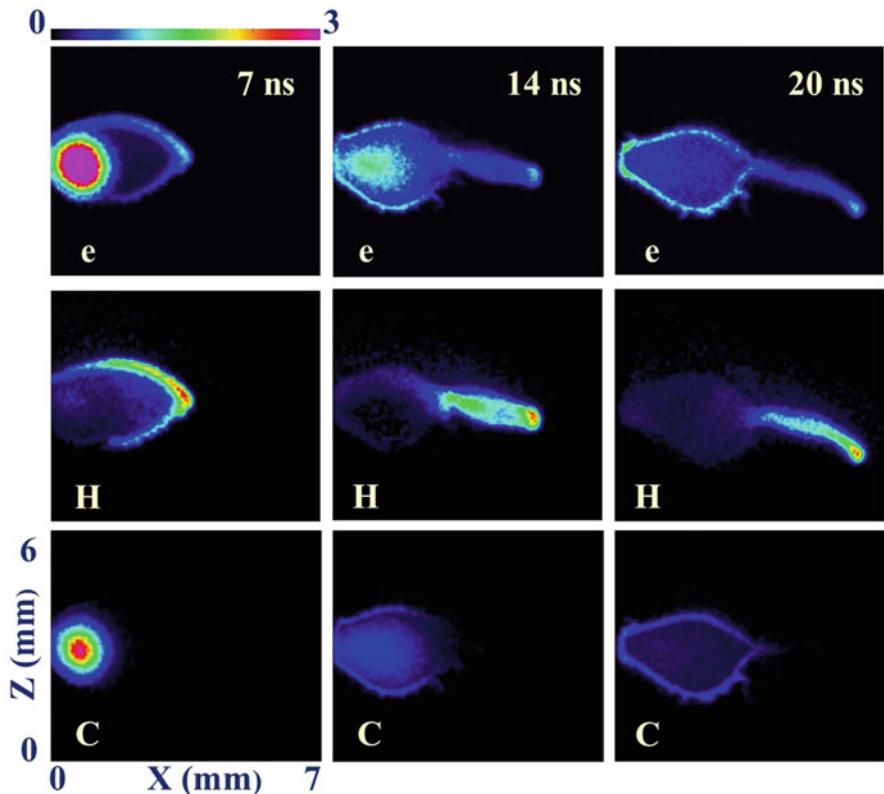


Fig. 13.11 Plasma density profiles obtained in a 3D HYPERS simulation of the UNR experiment. The CH plasma expands in a strong transverse magnetic field applied in the Y-direction. From bottom to top shown are: the carbon ion density, proton density, electron density. This 3D hybrid simulation reproduced the plasma features and scales observed in the experiment [33] and revealed a new ion species separation effect and explained the formation of a narrow collimated plasma flow

13.5.4 Plasma Turbulence

HYPERS is also set up to model externally driven and decaying plasma turbulence [34] at ion spatial scales. For instance, HYPERS was used to conduct 3D simulations of externally driven turbulence on a grid of $64 \times 64 \times 256$ ion inertial lengths ($256 \times 256 \times 512$ cells). These simulations [35] confirmed ion anisotropies, dynamic formation of current sheets, and magnetic reconnection in a turbulent plasma at MHD scales. In these simulations, EMAPS dynamically generated a range of local field timesteps spanning a range of two orders of magnitude. The magnetic spectra obtained in these simulations were found to be closely overlapping above ion gyro-scales with those in the corresponding full PIC simulations.

13.6 Prospective Applications

We have reported advances in multiscale plasma simulations enabled by a novel, EMAPS-driven massively parallel hybrid code, HYPERS, where field and particle updates are carried out as frequently as dictated by local physical laws and mesh geometry. The EMAPS formalism is general and may find adoption in a variety of scientific and engineering fields, where multiple timescales and nonlinear phenomena continue to present fundamental computational challenges.

On the one hand, large-scale PDE-based models with multiple timescales, discretized on irregular grids, require enormous resources in terms of CPU time when using traditional explicit time-stepping algorithms. On the other hand, the memory and parallel scalability constraints set size limits on physical problems that can be solved with implicit methods. Therefore, coupling self-adaptive discrete-event methods with matrix-free solvers on large irregular grids may enable accurate solutions of time-dependent hyperbolic and parabolic PDEs presently beyond reach. EMAPS offers a great potential for developing multiscale simulation codes in a number of different disciplines, including climate modeling (model coupling), geology (seismic events), medicine (neural computation), biology (heart modeling), fire propagation [36], and modeling of vacuum electronic devices [37]. For instance, event-driven algorithms have already been used to model plasma discharges [38]. The success of EMAPS has already spearheaded development of new solvers for oil and gas reservoir modeling (e.g., [39, 40]), where speedups of three orders of magnitude with respect to traditional time-stepping methods have been reported [39]. Below we briefly discuss some other prospective EMAPS applications.

13.6.1 Climate Modeling

Because of inevitable uncertainties in initial conditions and imperfect climate models, operational weather forecasts can be viewed only in a probabilistic sense. Large-scale climate and oil reservoir models are known to be very sensitive to the values of timesteps used. This often makes numerical convergence of such models difficult even in the absence of chaotic processes. Different timesteps may lead to largely varying quantitative predictions or even result in different regimes of the numerical solution. Using small timesteps results in simulation slowdowns and large timesteps lead to unacceptable errors or numerical instabilities. The end user in such situations may be unable to choose optimal timesteps for given conditions solely based on human intuition. EMAPS can help automate such scenarios. Specifically, an EMAPS-driven simulation is guaranteed to accurately follow physical variables in time with accuracy defined by a user and the underlying finite equations. Numerous plasma physics examples have proved superior convergence of event-driven simulations in change (δf) space compared to time-stepped simulations that use global timesteps as a convergence metric. Therefore, replacing time-driven simulations by event-driven analogs may lead to significantly shorter wall-clock

times for operational forecasts simultaneously with dramatic improvements in their accuracy.

13.6.2 Neuromorphic Computing

Event-driven computation is a biologically-inspired paradigm for representing simulation data as asynchronous events, similarly to neuron spikes in the brain. Recently the Brain-Inspired Computing group at IBM Research-Almaden has built the first gesture-recognition system implemented on event-based hardware. This IBM TrueNorth neurosynaptic processor generates on-demand power spikes, which makes it essentially an event-driven hardware analog of EMAPS. This type of hardware perfectly matches EMAPS-driven simulations which compute only “on-demand” changes. Therefore, this chip can provide computational power to event-driven simulations with efficient electric power usage. In other words, a self-adaptive EMAPS simulation, which computes only changes to the global state of a numerical system on an “as needed” basis, will be automatically consuming power “on-demand” when running on this massively parallel spiking neural network chip. Merging these two innovative technologies, inspired by Conway’s Game of Life, offers an enormous potential for revolutionizing computation as we know it today.

13.6.3 Neural Computation

Nearly all neuronal information processing and inter-neuronal communications in the brain involve action potentials, or spikes, which drive not only the short-term synaptic dynamics of neurons, but also their long-term dynamics. In many brain structures, action potential activity is considered to be sparse. This sparseness of activity has been exploited to reduce the computational cost of large-scale network simulations through the development of event-driven simulation schemes. However, existing event-driven simulations schemes use greatly simplified neuronal models, including those that use pre-calculated look-up tables to characterize synaptic and neuronal dynamics [41]. These event-driven simulations drastically differ from the EMAPS-driven simulations since they reduce the original underlying equations for action potential to “cellular automata” (CA), which perform calculations based on *predefined* neuronal properties. EMAPS does not require such simplifications since it solves the underlying multiscale equations through following adaptively computed changes that are always limited to within a specified physical “quantum” value. For simulations with spiking models and large synaptic connectivity, EMAPS may help resolve times of discrete action potentials in a high-fidelity manner, which cannot be currently achieved with time-stepped or CA models of large-scale neural networks.

13.6.4 Next-Generation Multi-Agent Systems

EMAPS can be readily extended beyond simulations of models based on physical laws or even evolution equations in general. A number of different technologies have been developed that utilize interacting agents. These include multi-agent system (MAS), agent-based model (ABM), and swarm intelligence. While there are differences in the details and objectives of these technologies, they share the same building blocks that consist of agents, agent–agent interactions, agent–environment interactions, and a parallel processing engine. One of the fascinating aspects of such systems is that a collection of agents, even when guided with simple rules, can exhibit intelligent and complex global behavior. This so-called collective intelligence is likely the path to the development of general artificial intelligence solutions, as opposed to using isolated very deep neural nets. The next logical step in the evolution of agent-based technology is to combine them with machine learning and reinforcement learning. EMAPS provides a foundational framework for development of this type of the next generation of agent-based systems. It has already demonstrated efficient parallelization on petascale computers and has built-in features for change detection and predictors. There are numerous applications for such multi-agent systems, including autonomous unmanned aerial vehicles (UAVs), distributed learning, blockchain, among others.

13.7 Conclusion

Many physical systems can be modeled by differential equations discretized in time and space. When these equations describe processes evolving on disparate timescales, in particular strongly coupled nonlinear gas, fluid and plasma phenomena, standard time integration techniques often have difficulty accurately and efficiently predicting dynamics of such systems. The root cause of these issues is the fact that all time-stepping methods update numerical solutions in a predefined time order. We have taken a different approach and abandoned the conventional *time-driven* integration in favor of a new *change-driven* methodology. EMAPS is an intelligent agent-based technology for modeling multiscale systems. The local agents in EMAPS are responsible for change predictions/detections and flux-conservative updates/corrections. The collective action of these agents is orchestrated by the scheduler agent (engine) which communicates with the local agents through event queues. Conceptually, EMAPS could be also viewed as a simulation time operating system (STOS). In this capacity it can serve as a universal platform for running multiscale models with built-in uncertainty and disparate timescales.

We have used the EMAPS technology to build a massively parallel, space-time adaptive hybrid code, HYPERS. In this chapter, we have demonstrated its capabilities on a number of challenging space and laboratory plasma problems, including some of the largest and best resolved hybrid simulations of the Earth's

magnetosphere. Finally, we have discussed broader impacts of the new technology on rapidly developing scientific and engineering fields.

Acknowledgments The authors are grateful to Vadim Roytershteyn for his assistance with magnetospheric simulation videos. Yuri Omelchenko acknowledges the NASA support of recent HYPERS developments (in particular, stretched mesh, parallel domain decomposition and boundary condition capabilities) under NASA Grant 80NSSC19K0838 at Trinum Research, Inc., and the helpful guidance of Li-Jen Chen in this work.

References

1. C. Rovelli, *The Order of Time* (Riverhead Books, 2018)
2. M.J. Berger, J. Oliker, J. Comp. Phys. **53**, 484 (1984). [https://doi.org/10.1016/0021-9991\(84\)90073-1](https://doi.org/10.1016/0021-9991(84)90073-1)
3. L. Krivodonova, J. Comp. Phys. **229**, 8537–8551 (2011). <https://doi.org/10.1016/j.jcp.2010.07.037>
4. Y.A. Omelchenko, H. Karimabadi, J. Comp. Phys. **216**, 179–194 (2006). <https://doi.org/10.1016/j.jcp.2005.12.008>
5. A. Lew, J.E. Marsden, M. Ortiz, M. West, Arch. Rational Mech. Anal. **167**, 85 (2003). <https://doi.org/10.1007/s00205-002-0212-y>
6. R. Abedi, S.-H. Chung, J. Erickson, Y. Fan, M. Garland, D. Guoy, R. Haber, J.M. Sullivan, S. Thite, Y. Zhou, in *Spacetime Meshing with Adaptive Refinement and Coarsening*. Proceedings of 20th Annual Symposium on Computational Geometry, Brooklyn, New York, USA (2004), pp. 300–308
7. J. Banks, J.S. Carson, B.L. Nelson, D.M. Nicol, *Discrete-Event System Simulation*, 3rd edn. (Prentice-Hall, Upper Saddle River, NJ, 2000)
8. R.M. Fujimoto, *Parallel and Distributed Simulation Systems* (Wiley Interscience, 2000)
9. H. Karimabadi, Y. Omelchenko, J. Driscoll, R. Fujimoto, K. Perumalla, D. Krauss-Varban, in *A New Simulation Technique for Study of Collisionless Shocks: Self-Adaptive Simulations*. AIP Conference Proceedings, vol. 781 (2005), pp. 56–63. <https://doi.org/10.1063/1.2032675>
10. J. Nutaro, B.P. Zeigler, R. Jammalamadaka, S. Akerkar, Lect. Notes Comp. Sci. **2660**, 319 (2003). https://doi.org/10.1007/3-540-44864-0_33
11. P. Valentini, T.E. Schwartzentruber, J. Comp. Phys. **228**, 8766–8778 (2009). <https://doi.org/10.1016/j.jcp.2009.08.026>
12. H. Karimabadi, J. Driscoll, Y.A. Omelchenko, N. Omidi, J. Comp. Phys. **205**, 755–775 (2005). <https://doi.org/10.1016/j.jcp.2004.12.003>
13. Y.A. Omelchenko, H. Karimabadi, J. Comp. Phys. **216**, 153–178 (2006). <https://doi.org/10.1016/j.jcp.2005.11.029>
14. Y.A. Omelchenko, H. Karimabadi, J. Comp. Phys. **226**, 282–300 (2007). <https://doi.org/10.1016/j.jcp.2007.04.010>
15. H. Karimabadi, J. Driscoll, J. Dave, Y. Omelchenko, K. Perumalla, R. Fujimoto, N. Omidi, Lect. Notes Comp. Sci. **3732**, 573–582 (2006). https://doi.org/10.1007/11558958_68
16. Y.A. Omelchenko, H. Karimabadi, J. Comp. Phys. **231**, 1766–1780 (2012). <https://doi.org/10.1016/j.jcp.2011.11.004>
17. T.R. Tang, K.S. Perumalla, R.M. Fujimoto, H. Karimabadi, J. Driscoll, Y.A. Omelchenko, SIMULATION, Tran. Soc. Model. Simulation Intern **82**, 61–73 (2006). <https://doi.org/10.1177/0037549706065481>
18. Y.A. Omelchenko, H. Karimabadi, in *Parallel Asynchronous Hybrid Simulations of Strongly Inhomogeneous Plasmas*. Proceedings of the 2014 Winter Simulation Conference (2015), pp. 3435–3446. <https://dl.acm.org/doi/10.5555/2693848.2694277>

19. Y.A. Omelchenko, L.-J. Chen, J. Ng, J. Geophys. Res. Space Phys. **126**, e2020JA029035 (2021). <https://doi.org/10.1029/2020JA029035>
20. D. Winske, L. Yin, N. Omidi, H. Karimabadi, K. Quest, Lect. Notes Phys. **615**, 136 (2003). https://doi.org/10.1007/3-540-36530-3_8
21. M. Palmroth, U. Ganse, Y. Pfau-Kempf, M. Battarbee, L. Turc, T. Brito, M. Grandin, S. Hoilijoki, A. Sandroos, S. von Alfthan, Living Rev. Comput. Astrophys. **4**, 1 (2018). <https://doi.org/10.1007/s41115-018-0003-2>
22. Y.A. Omelchenko, Phys. Rev. E **92**, 023105 (2015). <https://doi.org/10.1103/PhysRevE.92.023105>
23. M.E. Jones, D.S. Lemons, E.J. Mason, V.A. Thomas, D. Winske, J. Comp. Phys. **123**, 169–181 (1996). <https://doi.org/10.1006/jcph.1996.0014>
24. Y.A. Omelchenko, V. Roytershteyn, L.-J. Chen, J. Ng, H. Hietala, J. Atmos. Solar-Terr. Phys. **215**, 105581 (2021). <https://doi.org/10.1016/j.jastp.2021.105581>
25. Y.-H. Liu, J. Birn, W. Daughton, M. Hesse, K. Schindler, J. Geophys. Res. Space Physics **119**, 9773–9789 (2014). <https://doi.org/10.1002/2014JA020492>
26. Y.A. Omelchenko, T.E. Weber, R.J. Smith, in *3D Hybrid Simulations of Interactions of High-Velocity Plasmoids with Obstacles*. Proceedings of 57th Annual Meeting of the APS Division of Plasma Physics, Savannah, GA, November 16–20 (2015), Abstract BP12.011
27. H. Karimabadi, V. Roytershteyn, H.X. Vu, Y.A. Omelchenko, J. Scudder, W. Daughton, A. Dimmock, K. Nykyri, M. Wan, D. Sibeck, M. Tatineni, A. Majumdar, B. Loring, B. Geveci, Phys. Plasmas **21**, 062308 (2014). <https://doi.org/10.1063/1.4882875>
28. Y.A. Omelchenko, H. Karimabadi, H.X. Vu, in *Advances in Multiscale Simulations of Solar Wind Interactions with the Earth's Magnetosphere*. ASP Conference Series, Numerical Modeling of Space Plasma Flows, vol. 488 (2014), pp. 155–160
29. Y.A. Omelchenko, H. Karimabadi, Phys. Rev. Lett. **109**, 065004 (2012). <https://doi.org/10.1103/PhysRevLett.109.065004>
30. Y. Lin, X.Y. Wang, M.R. Brown, M.J. Schaffer, C.D. Cothran, Plasma Phys. Controll. Fusion **50**(7), 074012 (2008). <https://doi.org/10.1088/0741-3335/50/7/074012>
31. A. Collette, W. Gekelman, Phys. Plasmas **18**, 055705 (2011). <https://doi.org/10.1063/1.3567525>
32. Y.A. Omelchenko, in *Hybrid Simulations of the Penetration of Laser-Produced Plasmas Across a Magnetic Field*. Proceedings of 58th Annual Meeting of the APS Division of Plasma Physics, San Jose, CA, October 31–November 4 (2016), Abstract GP10.005
33. C. Plechaty, R. Presura, A.A. Esaulov, Phys. Rev. Lett. **111**, 185002 (2013). <https://doi.org/10.1103/PhysRevLett.111.185002>
34. S.S. Cerri, D. Grošelj, L. Franci, Front. Astron. Space Sci. **6**, 64 (2019). <https://doi.org/10.3389/fspas.2019.00064>
35. V. Roytershteyn, H. Karimabadi, Y. Omelchenko, K. Germaschewski, in *Kinetic Simulations of Collisionless Turbulence Across Scales*. Proceedings of Solar Heliospheric and Interplanetary Environment (SHINE 2016) Conference, Stowe, VT, July 5–10 (2015), id.117
36. I. Kaur, A. Mentrelli, F. Bosseur, J.-B. Filippi, G. Pagnini, Commun. Nonlinear Sci. Numer. Simulat. **90**, 300–320 (2016). <https://doi.org/10.1016/j.cnsns.2016.03.003>
37. D.-Y. Na, Y.A. Omelchenko, H. Moon, B.-H.V. Borges, F.L. Teixeira, J. Comp. Phys. **346**, 295–317 (2017). <https://doi.org/10.1016/j.jcp.2017.06.016>
38. T. Unfer, J. Comp. Phys. **236**, 229–246 (2013). <https://doi.org/10.1016/j.jcp.2012.11.018>
39. Q. Shao, S.K. Matthäi, L. Gross, J. Comp. Phys. **384**, 134–150 (2019). <https://doi.org/10.1016/j.jcp.2019.01.026>
40. D. Stone, S. Geiger, G.J. Lord, J. Comp. Phys. **342**, 161–176 (2017). <https://doi.org/10.1016/j.jcp.2017.04.026>
41. E. Ros, R. Carillo, E.M. Ortigoza, B. Barbour, R. Agis, Neural Comput. **18**, 2959–2993 (2006). <https://doi.org/10.1162/neco.2006.18.12.2959>