

DISCONTINUOUS GALERKIN MODELING OF WAVE
PROPAGATION, SCATTERING, AND NONLINEAR GROWTH IN
INHOMOGENEOUS PLASMAS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF
ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Forrest Foust
August 2012

© 2012 by Forrest Foust. All Rights Reserved.

Re-distributed by Stanford University under license with the author.



This work is licensed under a Creative Commons Attribution-Noncommercial 3.0 United States License.

<http://creativecommons.org/licenses/by-nc/3.0/us/>

This dissertation is online at: <http://purl.stanford.edu/qt890bb8059>

Forrest Foust

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.



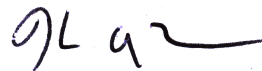
(Umran S. Inan) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.



(Maria Spasojevic)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.



(Howard Zebker)

Approved for the University Committee on Graduate Studies

Abstract

Modeling very low frequency (VLF, 3-30 kHz) wave propagation in the near-Earth space environment remains a significant computational challenge. Because of the strong inhomogeneity of the lower ionosphere, VLF waves propagating in these regions have an extremely wide range of wavelengths, varying from nearly zero to nearly infinite for some angles of propagation. In addition, the characteristic frequencies of a plasma may be an order of magnitude higher than the VLF frequencies of interest, presenting another significant problem for time-domain numerical solution methods because explicit time-stepping techniques are limited by the largest characteristic frequency present in the system. Further, the plasma environment of the Earth's magnetosphere is frequently unstable to VLF waves, leading to a wide variety of natural emissions and wave amplification phenomena, many of which are still poorly understood due to the relative paucity of high-resolution data and the difficulty in developing a complete theory of nonlinear wave growth in unstable plasmas.

Early approaches to modeling wave propagation in this environment relied heavily on analytical approximations under some set of physical assumptions, e.g., by linearization or by assuming the solution is smooth or slowly varying.

In recent years, supported by rapid increases in computer processing speed and memory capacity, so-called *continuum time-domain* methods such as the finite difference time-domain (FDTD) method have gained in popularity for solving such problems. FDTD works by discretizing a space into a finite number of points and then evolving the solution on these points forward in time, step by step. Although useful and simple, the FDTD method is not an ideal solution method for modeling wave propagation in the inhomogeneous ionosphere. FDTD is only accurate when both the

wavelengths present in the system are adequately sampled, and when the FDTD grid is highly regular (plaid) and slowly varying. The consequence of these restrictions is that when using FDTD, ionospheric propagation problems of any reasonable size are either severely underresolved in the short-wavelength regions of the space (meaning the solution is inaccurate), or overresolved in the long-wavelength regions of the space (meaning the memory requirements are too large).

Another continuum, time-domain solution technique, called the *discontinuous Galerkin* (DG) method, does not suffer from these restrictions. Due to its different formulation, the DG method is easily adapted for use on unstructured grids, allowing for increased spatial resolution only where it is required. The DG method is also highly accurate, yielding high-quality solutions with smaller memory and computational requirements than possible with low-order FDTD methods on structured grids.

This dissertation describes our approach to adapting the DG method for modeling wave propagation in plasmas, specifically addressing the challenges that arise at VLF in the near-Earth plasma environment.

We first present a procedure to incorporate any anisotropic linear dispersive material in the DG framework, with application to the perfectly matched layer (PML) and magnetized plasmas. As a semi-discrete formulation, we can exploit fully modern implicit-explicit Runge-Kutta time-stepping methods in order to circumvent the timestep restriction imposed by plasma frequency and gyrofrequency-induced stiffness, resulting in speedups of 5-10 times for typical mid-latitude ionospheres without the accuracy penalty typical of methods relying on low-order Strang splitting.

We next describe the hybrid DG-PIC (discontinuous Galerkin particle-in-cell) method for studying nonlinear wave growth and damping in magnetized plasmas, with specific attention paid to the scheme's efficient parallelization in distributed computing environments. Our DG-PIC method is a hybrid technique, which works by splitting the plasma into a cold, dense background plasma (modeled as a fluid) and a hot, strongly nonlinear and highly interacting population of energetic particles (modeled with superparticles). This model is far more efficient than a direct PIC method and also closely matches the physics of the real magnetosphere, making it an

ideal framework to investigate nonlinear wave phenomena there.

We demonstrate our techniques on two difficult problems of interest at VLF in the near-Earth space environment. First, we simulate scattering of VLF signals from intense lightning-induced ionospheric disturbances. VLF signals propagating in the Earth-ionosphere waveguide can be strongly scattered via interaction with density or temperature perturbations in the lower ionosphere, producing measurable perturbations on these signals when observed from ground-based receivers. To date, all simulations of this scattering process have relied heavily on approximations (such as smoothness or weak scattering) that do not, in general, hold. Our simulations represent, to our knowledge, the first direct simulations of scattering from such disturbances over a large 3D volume. The results have revealed the full spatial structure of the scattered field from intense lightning-induced disturbances as observed by a ground-based receiver.

Second, we use our hybrid DG-PIC technique to model the spontaneous generation of coherent rising emissions in an unstable, inhomogeneous magnetized plasma. The Earth’s magnetosphere naturally and frequently produces a variety of spontaneous emissions, driven by instabilities in the relatively energetic population of radiation belt particles that are injected into the Earth’s magnetosphere during active geomagnetic storms. One type of emission, termed *chorus*, is characterized by discrete, quasi-periodic, coherent rising tones. The nonlinear processes driving these emissions are, at present, still relatively poorly understood.

Our simulations demonstrate that it is possible to consistently and spontaneously generate quasi-periodic, coherent rising emissions, provided that the hot electron distribution driving the growth process is sufficiently unstable. The emissions are generated in a region just upstream of the magnetic equator and are subsequently amplified as they travel through the equatorial region and away from the generation region. We additionally show that many features of these rising emissions are well-predicted by linearized growth rates alone, demonstrating that while the chorus generation process is indeed nonlinear, there is nonetheless a close relationship between chorus generation and the linear growth processes that produce the high wave amplitudes necessary to induce nonlinear effects in the medium.

Acknowledgment

Thanks, everybody!

FORREST FOUST
Stanford, California
March 2012

This research was supported by a Stanford Graduate Fellowship, the NSF Office of Polar Programs award ANT-1141791, and the Office of Naval Research under Prime Award N00014-07-1-0789 to the University of Maryland with subcontract Z882802 to Stanford and under Grant N00014-09-1-0034-P00003.

Contents

Abstract	iv
Acknowledgment	vii
1 Introduction	1
1.1 Outline	2
1.2 Contributions	3
1.3 Notation	4
1.4 The ionosphere	6
1.5 VLF remote sensing of the lower ionosphere	8
1.6 The magnetosphere	10
1.7 Triggered emissions, chorus, and hiss	12
1.7.1 Chorus	13
1.7.2 Hiss	13
1.7.3 Triggered emissions	15
1.7.4 Other phenomena	16
1.8 Numerical modeling challenges	17
1.8.1 Length scales	17
1.8.2 Inhomogeneity	18
1.8.3 Plasma frequency-induced stiffness	18
1.8.4 Geometry	19
1.8.5 Nonlinearities	20
1.9 Survey of modeling techniques	20

1.9.1	Raytracing	21
1.9.2	Layered media	22
1.9.3	Mode-coupling approximations	23
1.9.4	Integral methods	23
1.9.5	Finite difference method	24
1.9.6	Finite volume method	25
1.9.7	Finite element method	26
1.9.8	The discontinuous Galerkin method	27
1.10	Strategies for modeling in plasmas	28
1.10.1	Linearizations	28
1.10.2	Particle-in-cell	29
1.10.3	Multi-moment fluid codes	30
1.10.4	Direct Vlasov simulation	30
1.11	History of numerical modeling	30
1.12	Commentary	35
2	Science background	36
2.1	Basic plasma theory	36
2.1.1	Definitions and single-particle motion	37
2.1.2	Distribution function	40
2.1.3	The Vlasov approximation	41
2.1.4	Linearization of the Vlasov equation	43
2.1.5	Moments of the Vlasov equation	44
2.1.6	Cold plasma approximation	46
2.1.7	The cold plasma conductivity tensor	47
2.2	Waves in plasmas	49
2.2.1	The cold plasma dispersion relation	49
2.2.2	The electrostatic hot plasma dispersion relation	51
2.2.3	The electromagnetic hot plasma dispersion relation	53
2.3	Plasma instabilities	55
2.3.1	Entropy arguments	56

2.3.2	Electrostatic beam instability	57
2.3.3	Whistler mode cyclotron instability	58
2.3.4	Inhomogeneity-driven cyclotron instability	59
3	The DG method	63
3.1	Description and derivation	64
3.1.1	Derivation	65
3.1.2	Nodal DG formulation	70
3.1.3	Application to Maxwell's equations	71
3.2	Linear dispersive media	73
3.2.1	Realization	76
3.2.2	Application to a cold plasma	79
3.2.3	Application to the PML	79
3.2.4	Rotations and transformations	84
3.3	Spectral properties of the DG scheme	85
3.4	Time integration and stiffness	91
3.4.1	Standard Runge-Kutta method	91
3.4.2	Low-storage Runge-Kutta method	92
3.4.3	IMEX Runge-Kutta method	93
3.4.4	Cold plasma IMEX method	94
3.4.5	Time integration convergence tests	96
3.5	Plasma verification	97
3.6	PML verification	102
3.7	Implementation notes	104
3.8	Summary	108
4	The DG-PIC method	110
4.1	PIC scheme	112
4.1.1	Projection step	114
4.1.2	Interpolation step	117
4.1.3	Divergence correction	119
4.1.4	Time integration	122

4.1.5	Hybrid approach	123
4.2	Validation	125
4.2.1	Grid heating	125
4.2.2	Conservation of momentum	126
4.2.3	Landau damping	128
4.2.4	Cold plasma	131
4.2.5	Cyclotron damping and growth	133
4.3	Summary	136
5	Ionospheric scattering	138
5.1	Background	138
5.2	Scattering model	140
5.2.1	Simulation domain	141
5.2.2	Plasma parameters	141
5.2.3	Incident source	142
5.3	Results	144
5.4	Summary	149
6	Simulation of spontaneous emissions	151
6.1	Background and prior work	152
6.2	Computational model	153
6.2.1	Mirror force	154
6.2.2	Particle loading	155
6.2.3	Particle sampling	156
6.2.4	Boundary conditions	159
6.3	Saturation and noise	159
6.3.1	Determination of linear saturation amplitudes	162
6.3.2	Discussion	164
6.4	Spontaneous emissions	165
6.4.1	Results	166
6.4.2	Spectral properties	168
6.5	Summary and commentary	171

7	Summary and Future Work	172
7.1	Adaptivity	173
7.2	DG Basis	173
7.3	Higher-order fluid moments	174
7.4	PIC	174
A	Inverse plasma dispersion relation	176
B	Derivation of the upwind flux	178
B.1	Two-characteristic derivation	178
B.2	Solution for the Maxwell system	180
B.3	Solution for the augmented Maxwell system	182
C	Runge-Kutta coefficients	184
C.1	Low-storage RK4	184
C.2	IMEX RK4	184
C.2.1	Explicit stage coefficients	185
C.2.2	Implicit stage coefficients	185
C.2.3	Common coefficients	185

List of Figures

1.1	Remote sensing with VLF	9
1.2	Example of an early-onset VLF scattering event	9
1.3	Magnetosphere and magnetospheric currents	11
1.4	Chorus emissions observed at Palmer	14
1.5	Hiss emissions observed at Palmer	14
1.6	Triggered emissions at Palmer	15
1.7	Mixed chorus and hiss emissions observed at Palmer	16
1.8	Unusual emissions observed at Palmer	16
1.9	Typical range of whistler mode wavelengths vs. altitude	18
1.10	Comparison between unstructured and structured meshes	20
1.11	Illustration of the PIC concept	29
1.12	Bush differential analyzer	31
1.13	EDSAC I in 1948	32
2.1	Forced single particle motion	38
2.2	Single particle phase space trajectory	41
2.3	Landau integration contour	52
2.4	Trapped and untrapped particle trajectories in phase space	61
3.1	Discretization of the continuous space	65
3.2	Conceptual illustration of Galerkin projection	67
3.3	Exploded view of DG elements and interpolation points	70
3.4	Node locations for polynomial orders 1, 3, and 5	71
3.5	PML illustration with stretching parameters indicated	81

3.6	Rotation of a gyrotropic medium	84
3.7	DG free space spectral characteristics	88
3.8	DG cold plasma spectral characteristics - well-resolved	89
3.9	DG cold plasma spectral characteristics - under-resolved	90
3.10	Stability boundaries for two explicit RK4 schemes	92
3.11	Explicit and IMEX time-stepping convergence	97
3.12	Convergence of the cold plasma DG scheme in 1D	98
3.13	Convergence of the cold plasma DG scheme in 2D	99
3.14	Theoretical cold plasma dispersion relation	100
3.15	Numerical dispersion relation in a cold plasma (central flux)	100
3.16	Numerical dispersion relation in a cold plasma (upwind flux)	101
3.17	PML convergence test for a rectangular domain	103
3.18	Illustration of the PML truncating a curved domain	104
3.19	PML convergence test for a curved domain	104
3.20	Grid segmentation	105
3.21	Local dense DG operations	106
3.22	CPU time per timestep vs. block length	107
3.23	Speedup vs. number of CPUs	108
4.1	Scattered point search problem	115
4.2	Cross-CPU point projection	117
4.3	Point containment ambiguity	119
4.4	Bounding box kd-tree illustration	119
4.5	Sparsity pattern for the DG operator matrix	123
4.6	Conservation of energy for the DG-PIC scheme	127
4.7	Conservation of momentum for the DG-PIC scheme	128
4.8	DG-PIC Landau damping verification	131
4.9	DG-PIC cold plasma dispersion relation - small particles	133
4.10	DG-PIC cold plasma dispersion relation - large particles	134
4.11	Hybrid DG-PIC cyclotron damping verification	136
5.1	Domain for the lightning-EMP scattering problem	142

5.2	Plasma height profiles and perturbations	143
5.3	Combined propagation and scattering model	143
5.4	Pure-mode scattered fields for vertical lightning EMP	145
5.5	Pure-mode scattered fields for horizontal lightning EMP	146
5.6	Vertical reference field for scattering problem	147
5.7	Full mode sum scattered fields - vertical lightning EMP	148
5.8	Full mode sum scattered fields - horizontal lightning EMP	149
6.1	Chorus simulation domain	154
6.2	Sample hot particle distribution function	155
6.3	Particle boundary conditions	159
6.4	Sample of linear saturation	163
6.5	Simulations of linear saturation	164
6.6	Chorus simulation	167
6.7	Linear growth rates	168
6.8	Linear growth rate and chorus spectrum comparison	170
B.1	Illustration of the two-characteristic Riemann problem	179

Chapter 1

Introduction

Modeling electromagnetic wave propagation in even simple media is an exceedingly difficult problem. Seemingly simple problems can be solved only approximately, necessitating the use of specialized techniques valid under certain limits or the application of approximations or linearizations to the governing equations before any attempt is made to solve the problem. In a magnetized plasma (a partially ionized gas composed of unbound electrons and ions), the situation is even more complicated and often completely intractable.

The focus of this dissertation is to describe the development of a new modeling framework based on a relatively new technique called the nodal discontinuous Galerkin (DG) method [[Hesthaven and Warburton, 2002](#)]. The DG method is highly parallelizable, has an arbitrary order of accuracy, and is easily adapted to unstructured grids. Our goal is to develop a generic platform for modeling a very large class of electromagnetic wave propagation problems in both the linear and nonlinear regimes. The near-Earth plasma environment presents a particularly challenging modeling problem due to the widespread presence of non-equilibrium distributions, very large length scales, slow propagating modes, and short characteristic timescales. We demonstrate the validity of our model in both linear and nonlinear regimes, showing that it is well suited to modeling two unrelated, but notoriously difficult problems in the near-Earth plasma environment.

1.1 Outline

In this introductory chapter, we give a description of the modeling environment of interest, the near-Earth plasma medium, including a brief overview of some natural radio phenomena at VLF (very low frequency, 3-30 kHz). We discuss some of the challenges for modeling wave propagation in this environment, followed by a survey of common numerical methods that can be applied to this problem. We then conclude with a historical discussion of the field of numerical simulation, showing how it paralleled and was influenced by the simultaneous development of the computer.

In Chapter 2, we provide the theoretical background for the remainder of the dissertation. We begin with an overview of the basic physics of plasmas. We begin with a microscopic view and generalize to the multi-moment fluid approach, followed by a discussion of waves in plasmas. We continue with a discussion of plasma instabilities from the point of view of non-equilibrium distributions, then provide some specific examples for which a linear growth rate can be derived.

In Chapter 3, we describe the DG method for modeling propagation in cold magnetized plasmas. We begin with a derivation of the DG method in three dimensions. We then describe our method for including dispersive currents in the DG framework, demonstrating its validity by applying it to two types of linear dispersive media: the perfectly matched layer (PML) and a multi-component cold plasma. We then follow with a discussion of the spectral properties of the DG operator with the cold plasma currents included, followed by a description of time-stepping methods, with particular emphasis placed on techniques for dealing with grid-induced and plasma frequency-induced stiffness. We conclude with a high-level description of our implementation, describing issues of parallelization, blocking strategies, and implicit solvers.

In Chapter 4, we describe our extension to the basic DG technique to include nonlinear effects via the DG particle-in-cell (DG-PIC) method. We discuss our extensions to the basic technique, paying careful attention to its efficient parallelization on unstructured grids. We conclude with a set of canonical experiments, demonstrating that the scheme reproduces physical behavior for cases in which an analytical solution is known.

In Chapter 5, we describe the application of the cold plasma DG method to scattering from lightning-induced ionospheric disturbances. Large lightning discharges are thought to strongly perturb the ionosphere directly above a lightning strike. This disturbance in turn perturbs the amplitude and phase of a narrowband transmitter signal propagating through the disturbed region, which can be detected and measured by a ground-based receiver. While the basic causal chain had already been well-established by experiment and prior two-dimensional models, the full three-dimensional structure of the scattered field induced by these disturbances was not known to a high degree of accuracy. This chapter addresses this issue, computing the full three-dimensional scattered fields from a lightning-induced ionospheric disturbance over a large ground area.

In Chapter 6, we use the DG-PIC method to model a notoriously difficult problem in magnetospheric plasma physics: spontaneous chorus emissions. Chorus is the name given to naturally-occurring radio emissions originating from the Earth's magnetosphere. The emissions are highly intense, coherent, and structured, often with characteristic discrete rising tones and quasi-periodic behavior. We apply our technique to modeling this phenomenon on a truncated one-dimensional domain, showing that we can consistently reproduce many of the qualitative features of chorus.

Finally, we make some concluding remarks in Chapter 7.

1.2 Contributions

The principal contributions of this work concern the development of new techniques for simulating wave propagation in strongly inhomogeneous, unstable plasmas. Much of this work concerns the details of these methods, but the final chapters concern the application of these methods to the solution of two problems of interest at VLF in the ionosphere and the near-Earth space environment.

Specifically, we have made the following contributions:

- We have developed techniques for the efficient simulation of Maxwell's equations in arbitrary, anisotropic dispersive media on unstructured grids using the DG method.
- We have applied these techniques to simulating both cold plasmas and the

perfectly matched layer, showing that the scheme maintains high-order accuracy and performs well even in the presence of abrupt material discontinuities.

- We have shown for VLF waves propagating in cold plasmas typical of the Earth's ionosphere, the maximum eigenvalue of the discretized system is bound by the upper right-hand mode cutoff frequency, not by the grid spacing. This introduces a severe timestep restriction when using explicit timestepping techniques. We have shown that with our formulation, we can use a high-order split Runge-Kutta method to circumvent this timestep restriction with no impact on the scheme's formal order of accuracy.
- We have used these methods to directly simulate VLF scattering from intense lightning-induced ionospheric disturbances. These simulations are, to our knowledge, the largest continuum time-domain simulations of this scattering process carried out to date.
- We have developed a large-scale hybrid particle-in-cell (PIC) code in the DG framework, developing new techniques to efficiently parallelize the scheme on unstructured meshes.
- Through parametric studies, we have demonstrated that the saturation amplitude for cyclotron wave growth in a uniform, unstable plasma is directly related to the linear growth rate itself. Since the saturation amplitude must be higher than any statistical fluctuations introduced by the finite particle sampling, these simulations conclusively show that long-time PIC simulations of cyclotron wave growth are only possible for very large linear growth rates.
- We have used the DG-PIC technique to model spontaneous emissions from unstable space plasmas. These simulation results qualitatively share many features with natural emissions and additionally predict that the spectra of these spontaneous emissions should be strongly correlated with linear growth rate spectra.

1.3 Notation

In this dissertation, we use the following notation. A simple scalar quantity is denoted in italics, e.g., *a*. Bold letters **a** refer to vector quantities. Whether it refers to a

spatially oriented vector, e.g., $\mathbf{a} = (a_x, a_y, a_z)$ or an N -dimensional vector, e.g., $\mathbf{a} = (a_1, a_2, \dots, a_N)$ depends on context. When it is convenient, we use shorthand notation \mathbf{a} to refer to a spatially oriented vector of vectors, e.g., $\mathbf{a} = (\mathbf{a}_x, \mathbf{a}_y, \mathbf{a}_z)$, where each component \mathbf{a}_i is itself a vector. Which meaning is intended is clear from context. A hatted quantity $\hat{\mathbf{a}}$ refers to a unit vector. The vector \mathbf{r} denotes the position vector, that is, $\mathbf{r} = (x, y, z)$ in three dimensions, $\mathbf{r} = (x, y)$ in two dimensions, and $\mathbf{r} = (x)$ in one dimension. In integrals, $\int d\mathbf{r}$ is shorthand to refer to the triple integral $\iiint dx dy dz$ in 3D, a double integral $\iint dx dy$ in 2D, or a single integral $\int dx$ in 1D. The same notation is used for the velocity vector $d\mathbf{v}$ and the momentum vector $d\mathbf{p}$.

Sans-serif characters such as \mathbf{A} refer to matrices or tensors. Which meaning is intended is clear from context. Due to typesetting limitations, Greek characters denoting tensor quantities are marked with a double arrow, e.g., $\vec{\sigma}$. The special matrix \mathbf{I} always refers to the unit dyad or identity matrix. For most of the derivations in this dissertation, we limit ourselves to only rank-two tensor fields, that is, the divergence of a tensor field $\nabla \cdot \mathbf{f}$ is a vector field:

$$\nabla \cdot \mathbf{f} = \frac{\partial \mathbf{f}_x}{\partial x} + \frac{\partial \mathbf{f}_y}{\partial y} + \frac{\partial \mathbf{f}_z}{\partial z}$$

Overloading notation in this manner is conceptually convenient when generalizing the DG scheme to tensor fluxes, but where required for clarity, we occasionally use indices, which allows us to clearly disambiguate, for instance, the inner and outer vector products, as illustrated in the following table:

$\mathbf{a} \cdot \mathbf{b} = a_i b_i$	(inner, or dot product)
$(\mathbf{ab})_{ij} = a_i b_j$	(outer, or dyadic product)
$(\mathbf{a} \times \mathbf{b})_i = \epsilon_{ijk} a_j b_k$	(cross product)
$\nabla \cdot \mathbf{u} = \frac{\partial}{\partial x_i} u_i$	(divergence of a vector field)
$(\nabla \cdot \mathbf{f})_j = \frac{\partial}{\partial x_i} f_{ij}$	(divergence of a tensor field)
$(\nabla \times \mathbf{u})_i = \epsilon_{ijk} \frac{\partial}{\partial x_j} u_k$	(curl)

The permutation symbol ϵ is defined as:

$$\epsilon_{ijk} = \begin{cases} 1 & \text{if } (i, j, k) = (1, 2, 3), (2, 3, 1), \text{ or } (3, 1, 2) \\ -1 & \text{if } (i, j, k) = (3, 2, 1), (2, 1, 3), \text{ or } (1, 3, 2) \\ 0 & \text{otherwise} \end{cases}$$

Note that we use the Einstein summation convention, where repeated indices are summed.

Throughout this text, we use j to denote the imaginary number $(-1)^{1/2}$. We use the complex frequency sign conventions $e^{-j\omega t}$ for time-domain phasors and e^{jkx} for the spatial domain.

1.4 The ionosphere

The *ionosphere* is the upper portion of the Earth's atmosphere, so named because it is partially ionized by the sun's radiation during the day. The absorption cross-section varies with altitude, but the usual photoionization sources are soft X-ray and extreme UV emissions [Kelley, 2009, p. 30]. Typical ionization densities are quite low – the ionized to neutral density ratio is on the order of 1% or less [Yeh and Liu, 1972, p. 402]. However, due to the low collision rates (decreasing exponentially with altitude), the ionosphere nonetheless behaves as a plasma, albeit a cold one, with temperatures on the order of a few thousand degrees Kelvin [Bilitza and Reinisch, 2008], relatively low for an ionized gas.

The practical consequences of the existence of the ionosphere are significant. Long range over-the-horizon radio communication is only possible because of the presence of the ionosphere. To radio waves in some range of frequencies, the lower boundary of the ionosphere acts as a reflective layer, making communication over very long distances possible. Very low frequency signals (VLF) in the range of 3-30 kHz, in particular, can propagate for extremely long distances with only minimal attenuation, effectively guided in the waveguide formed between the lower boundary of the ionosphere and the Earth [Davies, 1990, p. 367]. The ionosphere also affects higher frequency signals. High frequency (HF) signals, for instance, can have their ray paths bent by the

ionosphere. While not reflecting in the usual sense, this phenomenon can also be used for long-range communication. Global Positioning System (GPS) signals, while in the microwave range (in the GHz), nonetheless have their signals perturbed enough by the ionosphere that real-time, dynamic compensation using a system of ground-based receivers is necessary when high position accuracy is required [Enge *et al.*, 1996].

The ionosphere, however, is extremely variable, especially at night and at high latitudes, which presents a reliability problem for radio signals passing through or reflecting off of it. There are a number of reasons for this variability. At night, the ionosphere is no longer being actively ionized by the sun and thus relaxes, but some ionization sources still remain. Scattered Lyman-alpha radiation may contribute [Strobel *et al.*, 1974] as well as cosmic ray fluxes [Bailey, 1959]. Lightning discharges can also modify the ionosphere locally due to the enormous movement of charge, which creates large electric fields that can directly modify levels of ionization above a lightning strike [Pasko *et al.*, 1995, 1998]. In addition, the radiated EMP from a lightning stroke can have large enough electric fields to further modify the ionization above a lightning discharge [Inan *et al.*, 1991; Cheng *et al.*, 2007b; Marshall *et al.*, 2008a, 2010]. The intense radio waves from powerful VLF transmitters can also create strong electron density irregularities [Bell *et al.*, 2004, 2008; Titova *et al.*, 1984; James and Bell, 1987; Tanaka *et al.*, 1987] as well as heat the ionosphere over extended regions [Graf *et al.*, 2011].

Given its importance to long-range communication and to the advancement of scientific knowledge about atmospheric chemistry, it is remarkable that so little is known at present about the ionosphere's meso-scale structure. The primary reason is that direct measurements, particularly of the lower ionosphere, are extremely difficult due to its altitude (70-85 km), which is too low for satellites and too high for balloons. In addition, the electron density in the lower ionosphere is typically too low, especially at night, for there to be detectable echoes from radio sounders, even with the largest ground-based radars [Sechrist, 1974].

GPS signal delays, which can be related to the total electron content (TEC) between the satellite and the ground receiving site, can be used to perform rudimentary

tomographic reconstruction of the ionospheric density over both space and time [*Escudero et al.*, 2001; *Mitchell et al.*, 1997; *Brunini et al.*, 2004; *Tsai et al.*, 2002; *Pakula et al.*, 1995; *Ruffini et al.*, 1998; *Erturk et al.*, 2009]. However, these methods have fundamental resolution limits due to the sparsity of the measurements [*Raymond et al.*, 1994]. Further, imaging the lowest region of the ionosphere (between 60 and 90 km) is extremely difficult using tomographic methods due to the relatively small contribution of that region to the total electron content measured along a tomographic path [*Kunitsyn and Tereshchenko*, 2010, p. 88]. These limits have practical consequences. For instance, one of the more popular ionospheric models, the International Reference Ionosphere (IRI) is known to have systematic biases, problems reproducing actual electron densities during storm times, and inaccuracies on short spatial scales. These deficiencies have been noted by a number of authors [*Bittencourt and Chrysafidis*, 1994; *Jakowski et al.*, 1998; *Coisson et al.*, 2008; *Dettmering et al.*, 2011; *Lühr and Xiong*, 2010].

While direct measurements of the lower ionosphere remain difficult, some specialized tools to probe the structure of the lower ionosphere do exist, one example of which is so-called *VLF remote sensing*, utilizing propagation of VLF signals in the Earth-ionosphere waveguide.

1.5 VLF remote sensing of the lower ionosphere

Since VLF signals reflect near the lower boundary of the ionosphere between 70-85 km altitude, a VLF wave propagating in the Earth-ionosphere waveguide is exceptionally sensitive to ionospheric perturbations at those altitudes. Any low-altitude ionospheric density variation along the path from a transmitter to a receiver scatters the incident wave, resulting in a measurable perturbation on the amplitude and phase of the signal at the receiver. Since many powerful VLF transmitters are used worldwide for long-range navigation and communications, this technique can be an effective tool for measuring local disturbances in the lower ionosphere. A diagram of the basic process is illustrated in Figure 1.1.

In Figure 1.2, we show a typical example of a scattering event. The plot shows the magnitude of a filtered narrowband VLF transmitter signal as a function of time.

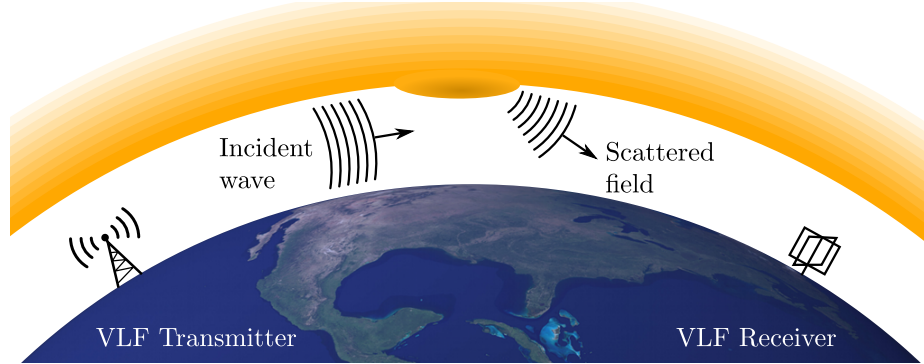


Figure 1.1: Illustration of the VLF remote sensing concept. An incident wave from a VLF transmitter at the left propagates in the Earth-ionosphere waveguide. The signal scatters from an ionospheric perturbation, which is then observed at the receiver.

At about 115 seconds, some localized portion of the ionosphere along the great circle path between the transmitter and receiver suddenly changed, manifested on the received signal as an abrupt change in the signal amplitude. In this case, the event is well correlated in space and time with a large lightning strike located along the transmitter-receiver path. The challenge faced by the modeler is to determine what physical mechanisms are consistent with such an observation.

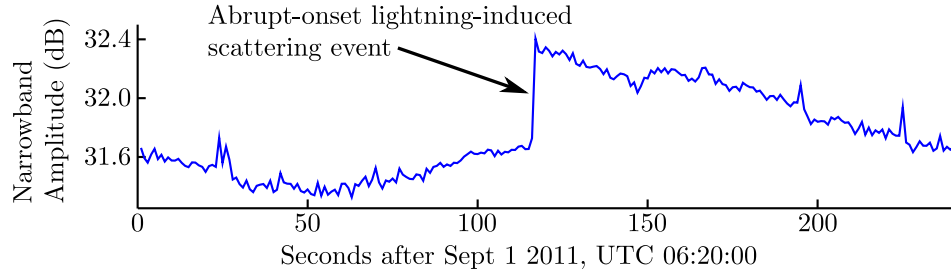


Figure 1.2: Example of an early-onset VLF scattering event. The plot shows the narrowband signal amplitude of the 24 kHz NAA VLF transmitter in Cutler, Maine as observed from a ground-based VLF receiver in Fountain, Colorado. A lightning-induced scattering event occurs at approximately 115 seconds, manifesting as an abrupt change in the received signal amplitude.

The use of VLF signals to probe the lower ionosphere is a well-established technique [Cheng *et al.*, 2007a; Moore *et al.*, 2003; Inan *et al.*, 1996; Cummer, 1997; Rodger and McCormick, 2006]. However, interpretation remains a problem due to the extreme sparsity of the data. In the absence of more dense receiver networks, modeling plays a crucial role in this process. Modeling work to determine the scattered signals from simple ionospheric disturbances has typically relied on analytical

approximations, for example, modal scattering approximations using Green’s function expansions [Poulsen *et al.*, 1990, 1993], mode-coupling calculations in cylindrical coordinates [Pappert and Snyder, 1972; Pappert and Ferguson, 1986], or Born scattering approximations in layered media [Lehtinen *et al.*, 2010]. Recently, direct simulation in 2D (and small 3D domains) has also become feasible [Marshall *et al.*, 2008a].

1.6 The magnetosphere

Processes determining the gross structure of the ionosphere in the range of 70-2000 km are largely independent of the Earth’s magnetic field. They are instead driven by bulk gas and plasma processes: internal pressure, gravity, molecular constituent densities, large-scale electric fields, and diffusion. The neutral constituents of the atmosphere tend to stratify into layers. The specific percentages of the molecular and atomic constituents as a function of altitude are largely determined by their masses. Lighter constituents have higher rates of diffusion and therefore dominate at higher altitudes [Hobbs, 2000, p. 29]. The structure of the ionosphere is directly related to this neutral structure, due to the different neutral constituent densities and differing corresponding absorption cross-sections to ionizing radiation. Above the ionosphere is a region known as the *magnetosphere*, where the Earth’s magnetic field begins to dominate the behavior of the ionized gases. The inner region of the magnetosphere, termed the *plasmasphere*, is composed primarily of cold, dense plasma, filling a magnetic field-aligned region surrounding the Earth out to approximately 2.5 to 6 equatorial Earth radii, dependent on ambient conditions. The ultimate source of this cold background plasma is diffusion from the ionosphere itself.

Beyond the plasmasphere, the Earth’s magnetic field can be strongly affected by the solar wind. The supersonic flow of particles from the sun distorts the Earth’s magnetic field into a characteristic comet-like shape, compressed on the sun side and extending into a long tail on the night side. The outer magnetosphere is highly dynamic, subject to large-scale fluctuations and instabilities. The basic picture of the solar wind “compressing” the Earth’s magnetic field, while intuitively satisfying, is overly simplistic. Any magnetic field must be associated with a real current, which the magnetosphere has in the form of large, connected current structures. The major

current systems of the magnetosphere are indicated in Figure 1.3.

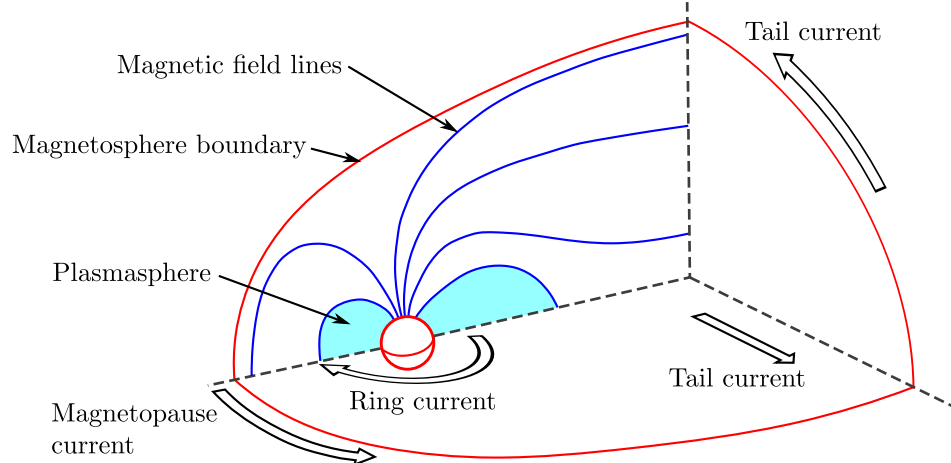


Figure 1.3: Illustration of the Earth’s magnetosphere, showing magnetic field lines, plasmasphere (shaded), and magnetopause (in red). Major magnetospheric currents are indicated with large arrows.

The magnetopause current, whose existence was originally postulated by [Chapman and Ferraro \[1931\]](#), is a large current structure formed at the boundary between the Earth’s magnetic field and the interplanetary magnetic field.

The sheet current [[Potemra, 1979](#); [Olson, 1984](#); [Antonova, 2004](#); [Baumjohann et al., 2010](#)] is a thin current structure flowing freely in the magnetic neutral region of the magnetotail, where the magnetic field lines on either side of the sheet connect to the North and South poles of the Earth. The sheet current and neutral region extend out to great distances on the tail side before eventually dissipating. The tail region of the magnetosphere can persist in metastable states in which large amounts of solar wind energy are stored and then rapidly released in *substorm events*, which are a result of sudden reconfigurations of the tail magnetic field of the tail [[McPherron, 1991](#)]. The process is closely related to magnetic reconnection, but the precise details are poorly understood at present. The increased magnetic reconnection rate during substorm events leads to enhanced transport and acceleration of tail plasma into the inner magnetosphere, where it joins and enhances the so-called *ring current* circling the Earth.

The ring current is largely a population of trapped particles in the inner magnetosphere existing in a magnetic field-aligned shell at approximately 3 to 5 equatorial

Earth radii. The trapped population of particles slowly drifts around the Earth, constituting a third large-scale current structure. During substorms, particles can be injected from the tail into the inner magnetosphere, replenishing this population of ring current particles. The electrons in this injected population have energies in the range of about 10 to 100 keV and have anisotropic velocity distributions. Such distributions are unstable and can amplify electromagnetic waves propagating through them [*Jordanova et al.*, 2010]. VLF whistler-mode waves in particular can resonate with the 10 to 100 keV electrons, gaining energy and being amplified in the process. Other sources can also be amplified, including simple thermal noise, natural radio emissions driven by lightning or other sources, or waves from ground-based transmitters, leading to a wide variety of wave phenomena, some examples of which we discuss in the next section.

1.7 Triggered emissions, chorus, and hiss

Any non-Maxwellian gas distribution is potentially unstable, a fact that can be deduced from a statistical mechanics point of view. For a given temperature, the Maxwellian distribution has the highest entropy of all possible distributions. Any non-equilibrium (non-Maxwellian) distribution, then, has free energy (energy that is available to do thermodynamic work) which could be converted to some other form. While this statement is factually correct, it is not especially useful since it says nothing about how one might actually extract the energy or what form it might take. However, it does give some useful insight into why, fundamentally, instabilities can occur.

The Earth's magnetosphere, and especially the region of the magnetosphere containing electrons with energies in the range of 10-100 keV superimposed on an otherwise Maxwellian cold plasma population, is a particularly interesting example of a non-equilibrium distribution. VLF whistler mode radio waves resonate with electrons in approximately this energy range. The unstable electron distribution gives up energy to the wave, amplifying it by as many as tens of decibels as the wave traverses the magnetosphere. The exact growth rates depend strongly on the wave frequency, the number of particles near resonance with the wave, and the degree of anisotropy (a

rough measure of the non-equilibriumness) of the distribution. This basic instability is natural and pervasive and is the primordial energy source for a number of natural VLF wave phenomena in the magnetosphere.

1.7.1 Chorus

Some of the most interesting consequences of magnetospheric wave amplification are nonlinear phenomena, encountered when wave amplitudes grow so high that the particle distribution not only gives up energy to the wave, but itself gets significantly perturbed by the wave. *Chorus* is one such example. Chorus refers to naturally-occurring, discrete VLF radio emissions spontaneously generated in the Earth's magnetosphere. The emissions usually take the form of discrete, quasi-periodic, intense rising tones, but can occasionally have falling tones, hooks, or even more complicated behavior. Chorus is often observed in two discrete frequency bands, an upper and a lower band, separated at approximately half the equatorial electron gyrofrequency [[Burtis and Helliwell, 1976](#)]. Experimental evidence suggests that chorus is generated outside the plasmasphere near the equator or near a magnetic minimum and is more common during times of high solar activity.

Chorus has been observed and studied for over fifty years, and while a number of theoretical generation mechanisms have been proposed, the specific details of the generation mechanism are still hotly contested. What is generally agreed upon is that chorus is caused by a nonlinear electron cyclotron wave-particle interaction driven by highly anisotropic electron distributions in the magnetosphere, and that the inhomogeneity of the background magnetic field plays some role in the development of the instability and the time-frequency behavior of individual chorus elements. A review paper by [Sazhin and Hayakawa \[1992\]](#) and more recently by [Santolík \[2008\]](#) summarize what is currently known about chorus, its generation mechanisms, and techniques for modeling. An example spectrogram, showing characteristic rising chorus elements, is shown in Figure [1.4](#).

1.7.2 Hiss

Hiss is the name given to another type of natural emission. As the name implies, hiss

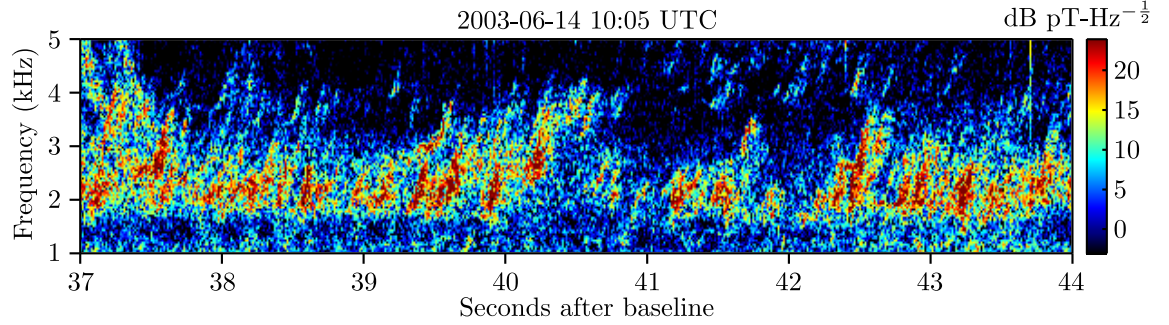


Figure 1.4: Magnetic field spectrogram of chorus emissions observed at Palmer station, showing a common type of emission with positive time-frequency ramps and quasi-periodic behavior.

is a diffuse, unstructured noise band with no discernible internal structure. Hiss is thought to be driven by similar mechanisms, i.e., amplification driven by anisotropic electron distributions in the magnetosphere. An example spectrogram recorded at Palmer Station, Antarctica, is shown in Figure 1.5. The literature on hiss is also extensive. A summary can be found in the review paper by *Hayakawa and Sazhin [1992]*.

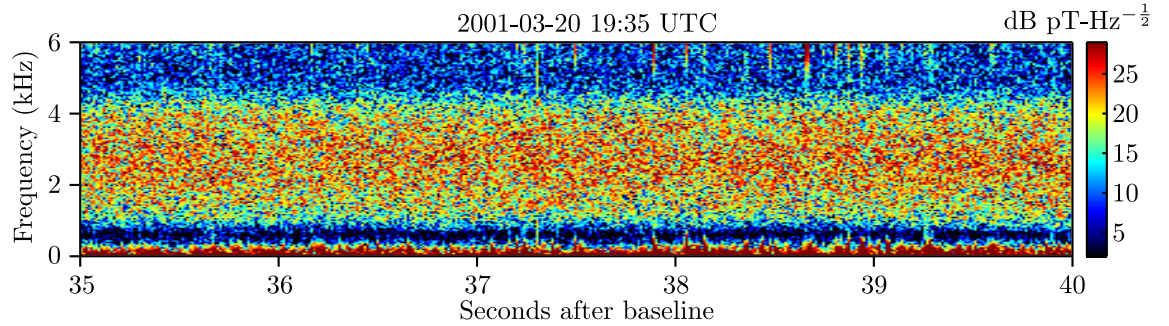


Figure 1.5: Magnetic field spectrogram of hiss emissions observed at Palmer station, showing a characteristic broadband noise band.

In contrast to chorus, hiss is broadband and diffuse with no obvious nonlinear character. Some authors have suggested that simple linear amplification of a broadband noise-like source may be sufficient to explain the amplitudes of hiss [*Solomon et al., 1988*]. Other work, however, suggests the opposite [*Church and Thorne, 1983*]. Precisely because hiss is so incoherent, resolving this disagreement is difficult. Tracing a hiss band to a single source region or even unambiguously determining the spectrum of wavenormal angles in hiss is extremely problematic.

A number of different potential embryonic sources of hiss have been proposed, by *Sonwalkar and Inan* [1989] and more recently by *Bortnik et al.* [2008, 2011]. In spite of the apparently contradictory views in the literature, however, it is plausible (even likely) that there are many different embryonic sources. VLF waves in the magnetosphere can follow complicated trajectories, drifting and magnetospherically reflecting for tens of seconds as they propagate. This fact, combined with the fact that a sufficiently unstable anisotropic electron distribution may amplify any broadband background noise to measurable levels means that the question of where a given packet of wave energy ultimately came from is largely irrelevant. Hiss is more likely to be observed when linear growth rates along all propagation paths from source to observation point are sufficiently large, and damping rates are sufficiently low, to make it observable. The characteristic incoherence may be due to incoherence in the seed waves or because of a lack of smoothness in the electron distribution itself, which would tend to destroy the coherent trapping structures necessary to support chorus-like growth.

1.7.3 Triggered emissions

A coherent VLF whistler mode wave passing through a region filled with a sufficiently anisotropic electron distribution whose electrons are in resonance with the wave is amplified. If the growth rate is sufficiently high, chorus-like emissions can be triggered by the wave. These emissions can be highly structured, showing both rising and falling features. An example is shown in Figure 1.6.

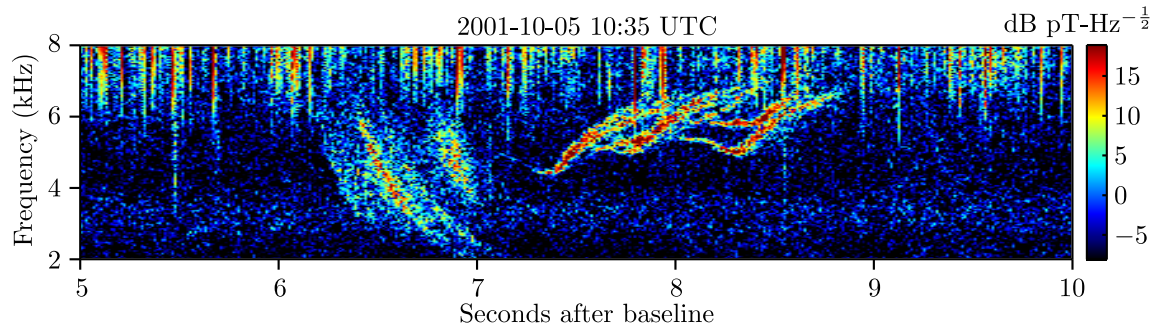


Figure 1.6: Magnetic field spectrogram of lightning-triggered emissions observed at Palmer station. The dispersed impulse from the causative lightning strike is at 6.5 seconds, triggering free-running emissions.

1.7.4 Other phenomena

Chorus and hiss are sometimes observed together, and hiss itself is sometimes seen to transition from a very diffuse, noiselike band into a more structured, chorus-like band and back. Occasionally discrete emissions are seen emerging from the top of a strong hiss band as shown in Figure 1.7, suggesting that hiss bands can sometimes mask underlying coherent emissions and that the same instability that amplifies hiss may also give rise to coherent emissions.

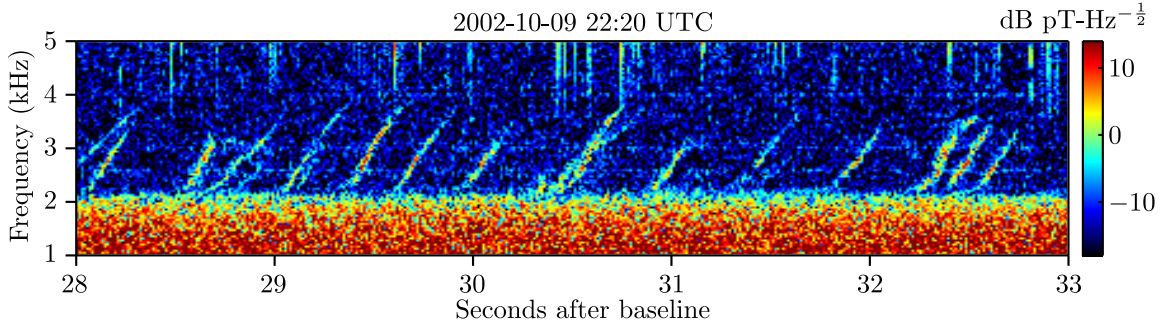


Figure 1.7: Magnetic field spectrogram of mixed chorus and hiss emissions observed at Palmer station. Quasi-periodic rising emissions are shown emerging from a hiss band.

Other emissions defy classification, sharing features in common with triggered emissions and with free-running, intense chorus emissions. One such example is shown in Figure 1.8. It is possible that these waves are triggered as well, but the causative source is too low in amplitude to be seen.

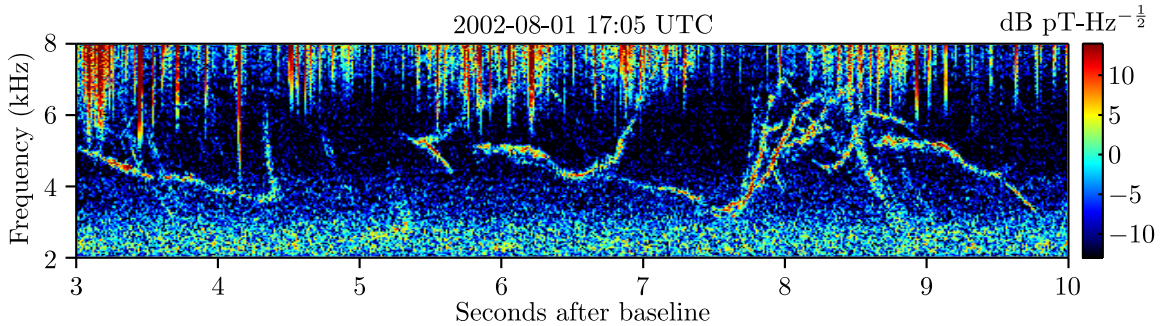


Figure 1.8: Magnetic field spectrogram of unusual free-running emissions observed at Palmer station.

1.8 Numerical modeling challenges

VLF wave propagation in the ionosphere and inner magnetosphere presents a difficult modeling challenge. Because of the large length scales of interest, extremely long timescales (on the order of seconds or more), the strong inhomogeneity in the electron density naturally present in the ionosphere, and the pervasive presence of nonlinearities, many significant problems of interest are completely intractable to solve even on modern supercomputers at the present time. The challenge a modeler faces, then, is to simplify the global problem into one that is tractable, which necessarily involves making approximations or reducing the size or scope of the simulation domain.

1.8.1 Length scales

The first major modeling challenges are due to the large length scales involved. VLF waves propagating in the Earth-ionosphere waveguide are only minimally attenuated, with attenuation rates of only a few decibels per megameter [Davies, 1990, p. 367]. Since VLF waves in the range of 3-30 kHz have free-space wavelengths of 10-100 km, typical problem sizes of interest have on the order of tens to hundreds of wavelengths per dimension. For numerical techniques that use spatial subdivision of the domain, this limits the size of a domain that can be modeled. In the ionosphere, this issue is exacerbated by the strong inhomogeneity of the ambient medium itself.

A magnetized plasma is both an anisotropic (the wavelength varies with propagation direction) and a dispersive medium (the wavelength varies with frequency). The two primary physical parameters affecting propagation at VLF are the electron gyrofrequency (Equation 2.4) and the plasma frequency (Equation 2.2). The gyrofrequency scales with the background magnetic field magnitude and the plasma frequency varies with the square root of the electron number density. The Earth's magnetic field is not a strong function of altitude near the Earth's surface, but its direction and magnitude do vary significantly as a function of latitude and longitude.

1.8.2 Inhomogeneity

The plasma frequency introduces problems when modeling sub-ionospheric propagation because of its strong height dependence, which in turn leads to strong height-dependent variation in the refractive index of VLF waves. As shown in Figure 1.9, for a fixed VLF frequency of 30 kHz, the wavelength can vary by over an order of magnitude, from 10 km in free space down to a minimum of less than a kilometer above approximately 100 km altitude. The strong spatial gradients, as well as the wide range in wavelengths even at a single, fixed frequency once again presents a significant problem for modeling. For techniques that rely on spatial subdivision, the problem is essentially one of sampling – the space must be subdivided more finely in areas of strong gradients and where the wavelength becomes small. The ideal case is to sample the space using an unstructured mesh, placing more grid cells only where they are needed. In this way, the extra resolution is not wasted in areas where the medium or wavelength are not changing rapidly.

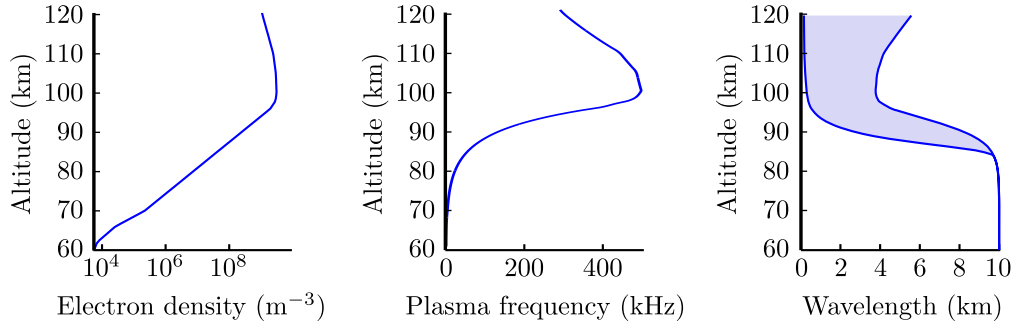


Figure 1.9: Electron number density, plasma frequency, and total range of allowable whistler-mode wavelengths (shaded region) for a 30 kHz VLF wave, plotted as a function of altitude. Parameters are typical of the mid-latitude ionosphere. The minima and maxima are determined by finding the minimum and maximum real wavelengths over all possible propagation angles for the right-hand circularly polarized mode.

1.8.3 Plasma frequency-induced stiffness

Another significant problem at VLF is that the characteristic frequencies of the plasma (the gyrofrequency and plasma frequency) can be an order of magnitude larger than the frequencies of interest (see Figure 1.9). This difference in scale affects time-domain solvers in the following way. Suppose that we wish to integrate (in time) a linear

ordinary differential equation (ODE):

$$\frac{d\mathbf{u}}{dt} = f(\mathbf{u})$$

Since the ODE is linear, we can consider each mode of the system separately by computing an eigenvalue decomposition and doing a coordinate transformation $\mathbf{u} \rightarrow \mathbf{v}$, where \mathbf{v} and ω are an eigenvector and eigenvalue, respectively, of the system $f(\mathbf{u})$:

$$\frac{d\mathbf{v}}{dt} = \omega \mathbf{v}$$

We would ideally like to solve this ODE using an explicit numerical method because explicit methods are efficient and easily parallelized. An explicit numerical method only uses information from the past to extrapolate the solution at some future timestep. The maximum timestep that we can take using an explicit method is limited by the magnitude of the largest eigenvalue ω_{\max} . In a plasma, ω_{\max} is related to the plasma frequency and gyrofrequency and can be as much as two orders of magnitude higher than the VLF frequencies of interest. Therefore, the maximum timestep we can take is less than an order of magnitude smaller than that required to get an accurate solution of the equation, and thus the total computational cost is an order of magnitude higher.

1.8.4 Geometry

A final challenge is geometry, that is, common computational problems involve curved or complicated structures or boundaries. Many numerical methods are easier to derive and implement on rectilinear (plaid) grids, but such frameworks place unnecessary restrictions on the types of problems that one might solve. For instance, a rectilinear grid cannot conform to a curved boundary (e.g., the Earth) without great loss of precision. One common technique to overcome this problem is to use a curvilinear coordinate system, but such a system imposes a similar set of constraints on the geometry, e.g., a curved coordinate system cannot conform to rectilinear boundaries. Unstructured meshes solve this problem by allowing the modeler to choose an arbitrary mesh to conform to the specific geometry of their problem, as illustrated

in Figure 1.10. Popular unstructured mesh techniques for electromagnetics, however, are either restricted to low order (low accuracy), as in finite difference techniques, or involve expensive matrix operations, as in finite element techniques.

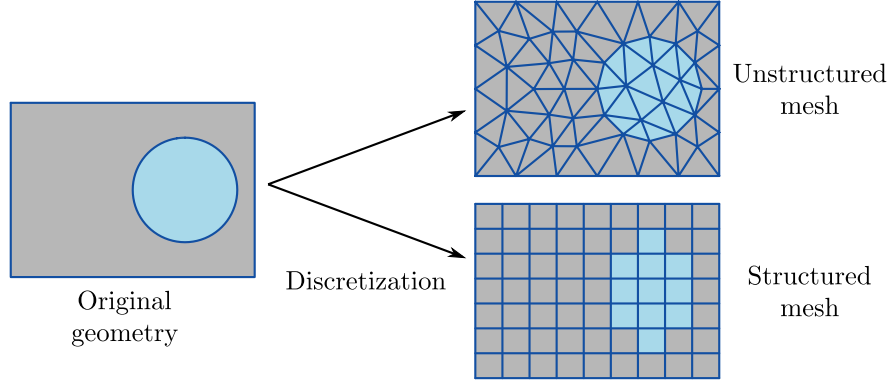


Figure 1.10: Comparison between unstructured, conformal meshes and unstructured meshes. The original geometry is at the left. Colors denote different types of materials. Even though the total number of mesh elements are similar in the two discretizations, the unstructured grid does a much better job of representing the curved boundary.

1.8.5 Nonlinearities

Nonlinear effects in the magnetized near-Earth plasma environment are commonplace and non-equilibrium electron distributions are the norm rather than the exception. The most complete fluid description of a plasma, which captures all linear and nonlinear effects, is a six dimensional nonlinear partial differential equation (PDE). Direct simulation of the complete six-dimensional system is unfortunately completely intractable at the present time except for small toy problems. Nonlinearities, therefore, are typically dealt with in problem-specific ways. Some computational strategies for dealing with nonlinearities in plasmas are discussed in Sections 1.10.3, 1.10.2, and 1.10.4.

1.9 Survey of modeling techniques for electromagnetics

This section provides an overview of some approaches commonly used to model wave propagation in the ionosphere and magnetosphere. Each technique ultimately aims

to solve Maxwell's equations for a magnetized plasma, but each has its own range of validity and its own class of problems for which it is well-suited to solve.

1.9.1 Raytracing

The raytracing approximation treats Maxwell's equations in only a very approximate sense, tracing the flow of a narrowband wave packet through a system. The technique is similar to the geometric optics approximation commonly employed to analyze lens systems, but generalized to spatially-varying media. The technique works by treating the dispersion relation as a conserved quantity (a Hamiltonian) of the system. The dispersion relation relates the wavenormal vector \mathbf{k} to the temporal frequency ω via some implicit function $F(\mathbf{k}, \omega, \mathbf{r}) = 0$ (see Section 2.2.1). By identifying the function F as the Hamiltonian, the wavenormal \mathbf{k} as a momentum, and ω as an energy in some reference frame, we can (by using Hamiltonian mechanics) directly write down a set of equations describing the path of a ray through an inhomogeneous medium [*Haselgrove, 1955*].

$$\begin{aligned}\frac{\partial \mathbf{r}}{\partial t} &= -\frac{\nabla_{\mathbf{k}} F}{\partial F / \partial \omega} \\ \frac{\partial \mathbf{k}}{\partial t} &= \frac{\nabla_{\mathbf{r}} F}{\partial F / \partial \omega}\end{aligned}$$

A more rigorous derivation is discussed by *Poevverlein* [1962]. The technique can be generalized to track the ray amplitude, also including higher-order effects like ray spreading [*Henye, 1980*].

The ray tracing approximation fails in the presence of strong gradients in the material parameters. Roughly speaking, the refractive index in the medium cannot change significantly over one wavelength. The basic scheme as described by *Haselgrove* [1955] also contains no information about the amplitude of the wave, nor can it handle diffraction effects. Amplitude information can be inferred by launching ray bundles, propagating some distance and then measuring the spatial density of the rays. In a similar manner, temporal behavior can be inferred by launching many rays at different frequencies and using Fourier decomposition methods.

The technique cannot be used in the lower ionosphere at VLF because of the high gradients in the refractive index, although it has some use in computing HF ray paths through the ionosphere. Ray tracing is primarily used when the length scales or time scales are far too large to use any other techniques, for example, in the case of modeling the slow propagation of whistler mode waves through the magnetosphere.

1.9.2 Layered media

The layered media approximation or *full wave method* [[Nagano et al., 1994](#); [Yagitani et al., 1994](#); [Nygren, 1981](#); [Wait, 1996](#)] is a steady-state technique used to find the solution to Maxwell’s equations in a stratified domain composed of horizontal uniform layers, infinite in extent. At each layer boundary and for each incident wavenormal of interest, Snell’s law is solved to find the two transmitted and reflected mode directions. Since the component of the wavenormal parallel to the surface is conserved on either side of a reflecting boundary (due to phase continuity), the problem amounts to solving for the component of \mathbf{k} normal to the surface and the associated modes (the electric and magnetic polarizations associated with those wavenormals).

These reflected mode directions and polarizations are then used to compute the reflection and transmission coefficients, which relate the incident mode amplitude to the two reflected and the two transmitted mode amplitudes. The system is then solved as a large matrix equation subject to boundary conditions to find these mode amplitudes for every wavenormal angle present in the system. Sources can be handled by transforming current distributions from physical space into k -space (i.e., using the Fourier transform). A recent technique published by [Lehtinen and Inan \[2008\]](#) solves the system using an efficient recursion relation and can also include any arbitrary distribution of source currents.

The method’s primary strength is its efficiency and ease of parallelization. Since the technique works in k -space, systems of great physical extent can be simulated to high precision with relative ease, provided the k -space spectrum is sufficiently sparse. Problems with very dense k -space spectra, however, would not benefit from this advantage. Additionally, the technique is not useful if there is lateral variation in the material properties. First-order perturbation techniques can be used to approximately

solve such problems but only if the variation is very small relative to the background [[Lehtinen et al., 2010](#)].

The layered media technique remains popular for modeling sub-ionospheric propagation over limited lateral extent, but it requires mode-coupling approximations or other similar methods in order to be used over longer distances.

1.9.3 Mode-coupling approximations

For propagation in the Earth-ionosphere waveguide, the limitations of the layered media method can be partially overcome using a mode-coupling approximation [[Pappert and Snyder, 1972](#); [Pappert and Ferguson, 1986](#)]. A long sub-ionospheric propagation path is divided into a finite set of segments. In each segment, a full-wave method or some other method is used to compute a set of waveguide modes and mode amplitudes. A set of mode coupling coefficients are then computed, which relate the modes and amplitudes in one slab to the modes and amplitudes in the next. Weak or evanescent modes are discarded. Backscattered modes are usually neglected [[Pappert and Snyder, 1972](#)]. This family of techniques is limited to propagation in waveguides and may have little use for any other type of modeling problem.

1.9.4 Integral methods

Integral methods are used to solve problems for which a *Green's function*, or integral kernel, can be found. A Green's function is essentially a statement of action at a distance. For instance, we can write the electric field at some distance \mathbf{r} as the sum of contributions from a set of currents $\mathbf{J}(\mathbf{r})$ distributed over space (presumably some small subset of the entire space). The time-harmonic electric field is given by an integral over those currents weighted by some distance-dependent kernel [[Harrington, 1961](#), pp. 124-125]:

$$\mathbf{E}(\mathbf{r}) = j\omega\mu \int \mathbf{G}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}(\mathbf{r}') d\mathbf{r}'$$

The kernel $\mathbf{G}(\mathbf{r}, \mathbf{r}')$ is a tensor Green's function that encapsulates how a current at some location \mathbf{r}' contributes to the electric field at some other location \mathbf{r} .

Green's functions are exceptionally difficult to derive for the arbitrary, inhomogeneous case. Therefore, integral methods are typically used only to solve simple scattering or radiation problems, where the problem is to find the fields at some distance from a small scatterer or antenna in an otherwise simple, uniform medium.

A further complication is that the source distribution $\mathbf{J}(\mathbf{r})$ is rarely known in advance for any practical problem. What is more typical is that the electric fields at some set of N points are known or prescribed (e.g., at a boundary). To determine the unknown current distribution $\mathbf{J}(\mathbf{r})$, we first expand it in some basis with N unknown expansion coefficients \mathbf{J}_i . Using the basis allows us to convert the integral into a simple weighted sum of the unknown expansion coefficients \mathbf{J}_i at some observation point \mathbf{r}_j . This procedure is repeated for each of the N observation points \mathbf{r}_j where the electric fields $\mathbf{E}(\mathbf{r}_j)$ are known or prescribed, each time inserting the resulting vector of weights into a new row in a matrix \mathbf{M} . Solving the system $\mathbf{M}\mathbf{J} = \mathbf{E}$ then yields all of the unknown expansion coefficients \mathbf{J}_i , effectively inverting the integral and solving the problem. This technique is termed the *method of moments* [Gibson, 2008; Sadiku, 2000]. The matrices involved are typically very large and dense, so this method is only practical when the number of unknowns N are relatively small.

1.9.5 Finite difference method

Finite difference methods [Lax and Keller, 1951; Lax, 1958; Lax and Wendroff, 1964, republished 2005, 1960; Courant et al., 1952; Weinberger, 1959; Yee, 1966; Taflov and Hagness, 2005; Inan and Marshall, 2011] are used to solve partial differential equations (PDEs) by dividing the space up into a set of discrete sample points (x_1, x_2, \dots, x_N) , and then approximating the continuous derivatives in the original system using finite differences, for example:

$$\left. \frac{\partial u}{\partial x} \right|_{x=x_i} \simeq \frac{u(x_{i+1}) - u(x_i)}{x_{i+1} - x_i} = \frac{u_{i+1} - u_i}{x_{i+1} - x_i},$$

where we have replaced the continuous unknown $u(x)$ with a finite set of unknowns (u_1, u_2, \dots, u_N) defined at the sample points (x_1, x_2, \dots, x_N) . Thus, we have converted the continuous equation into a discrete approximation, turning the PDE into a large

ODE. The ODE can then be solved using any time-stepping method.

The first application of the finite difference method to Maxwell's equations was described by [Yee \[1966\]](#). In recent years, the method has become the method of choice for solving difficult electromagnetics problems due to its ease of implementation. The finite difference method, unfortunately, suffers from a number of drawbacks that limit its usefulness for certain types of problems in electromagnetics, most notably: its poor handling of sharp material discontinuities, its poor handling of anisotropic materials, its low order of accuracy except on very smooth domains, and its difficult adaptation to unstructured grids while preserving high-order accuracy.

The fundamental issue causing all of these problems is the non-locality of the finite difference operator. Continuous derivatives are strictly local statements about the slope of a function, and the finite difference approximation violates this principle by incorporating non-local information to estimate the derivative at some given point.

1.9.6 Finite volume method

The finite volume method [[LeVeque, 2002](#)] is used to solve conservation law problems of the form:

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

where \mathbf{f} is a *flux* that describes how material flows into or out of a small control volume. The first step is to divide the space into a number of finite volumes of some arbitrary shape. In each volume i , the quantity $u(x)$ is first replaced by a single cell-averaged quantity \bar{u}_i . The method can be implemented as a numerical scheme by recognizing that the integral of the flux over some volume is related to the surface integral of the flux around that volume through Gauss's law:

$$\int_V \nabla \cdot \mathbf{f}(u) = \int_S \mathbf{f}(u) \cdot d\mathbf{S}$$

The fluxes at the surfaces between elements are estimated by combining the fluxes $\mathbf{f}(\bar{u})$ interior and exterior to an element in a physically-consistent way. While similar to the finite difference method, the different formulation makes the finite volume

method much easier to use on unstructured grids.

However, there are fundamental difficulties extending the finite volume scheme to high order accuracy. Essentially, the scheme only permits either high accuracy or unstructured grids. The extension to both high order and unstructured meshes is extremely problematic. In addition, the unique odd-even structure and skew symmetry of Maxwell's equations make the finite volume scheme extremely inaccurate (only first order) when naively applied to it. Complicated schemes, such as using interleaved grids for the electric and magnetic fields, are necessary to bring the finite volume scheme up to second-order accuracy for Maxwell's equations [*Madsen and Ziolkowski, 1990*].

1.9.7 Finite element method

The finite element method [*Jin, 2002; Hughes, 2000; Strang and Fix, 2008*] is another technique to discretize the spatial component of a PDE. The basic idea is to subdivide the space into a number of small subdomains, termed *finite elements*, and find some approximate solution in a basis expansion over each of those elements:

$$u(x) \simeq \sum_{j=1}^N u_j \phi_j(x)$$

The most straightforward way to convert this idea into a numerical scheme is by *Galerkin projection* [*Galerkin, 1915*]. For example, to solve the following PDE,

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = 0,$$

we substitute the approximation for $u(x)$, leaving a residual error $e(x)$:

$$\sum_j \frac{\partial u_j}{\partial t} \phi_j(x) + \sum_j f_j \frac{\partial \phi_j}{\partial x} = e(x)$$

We then multiply by every function in the basis and integrate over all space, setting the integrated residual to zero:

$$\sum_j \int \left(\frac{\partial u_j}{\partial t} \phi_j(x) + f_j \frac{\partial \phi_j}{\partial x} \right) \phi_i(x) dx = 0$$

Finally, we enforce continuity between each element and evaluate the integrals for every pair (i, j) in the basis and every element in the domain, yielding a very large matrix ODE:

$$\mathbf{M} \frac{d\mathbf{u}}{dt} + \mathbf{S}\mathbf{f} = 0$$

Once again, we have converted a PDE into an ODE with a large number of unknowns. The finite element method, at first glance, has many advantages. It is highly accurate, flexible, and simple to use on unstructured grids. The biggest disadvantage for time-domain problems is that the scheme is not local, that is, imposing the continuity constraint between elements means that the matrix \mathbf{M} is as large as the entire space. Solving this system of equations is extremely expensive for large problems, often completely dominating the computational cost. It is for this reason that finite element methods are not often used for time domain problems.

1.9.8 The discontinuous Galerkin method

The discontinuous Galerkin (DG) method, the subject of this dissertation, combines elements of the finite volume method and finite element method to create a scheme that is both local and highly accurate. The scheme also solves conservation law problems of the form:

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

but does so in a slightly different manner than the finite volume method, using a local basis expansion in each element instead of a volume-averaged quantity. The technique is somewhat less generic than the finite difference method since it can only be used for conservation law problems.

This method and its application to Maxwell's equations is discussed in far more detail in Chapter 3.

1.10 Strategies for modeling in plasmas

The Vlasov approximation, described in detail in Section 2.1.3, describes the behavior of a collisionless plasma in the fluid limit. It self-consistently describes the time evolution of a particle distribution function $f(\mathbf{r}, \mathbf{v}, t)$ in the presence of electric and magnetic fields \mathbf{E} and \mathbf{B} . Unfortunately, since the Vlasov equation is six dimensional and nonlinear, direct simulation is intractable. We now briefly discuss some of the common approximations used to simulate plasmas. In principle, these techniques are independent of the specific numerical method used.

1.10.1 Linearizations

One common approach is to linearize the Vlasov equation, that is, to assume that all of the fields and the distribution function itself consist of a zeroth-order term plus a small first-order perturbation.

$$f(\mathbf{v}) = f_0(\mathbf{v}) + f_1(\mathbf{v})$$

$$\mathbf{E} = \mathbf{E}_0 + \mathbf{E}_1$$

$$\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_1$$

For wave propagation problems, it is common to assume the zeroth-order electric field $E_0 = 0$, in which case the first-order field perturbations are the wave electric and magnetic fields, i.e., $\mathbf{E}_1 = \mathbf{E}_w$ and $\mathbf{B}_1 = \mathbf{B}_w$. The linearized Vlasov equation can then be further manipulated to yield dispersion relations for waves in a hot plasma, giving both the allowed wave modes and their respective growth or damping rates [*Inan and Golkowski, 2010*, Chapter 12]. The technique is primarily useful as an analytical tool, but the basic idea can also be used for simulation, leading to so-called δf techniques [*Dipesso et al., 1991*].

The assumption of small perturbations on a zeroth-order quantity can give powerful insight into the behavior of a plasma at small amplitudes but fails when the

first-order perturbation becomes large relative to the zeroth-order terms.

1.10.2 Particle-in-cell

The basic particle-in-cell (PIC) method [Birdsall and Langdon, 1985] models a plasma in a more direct fashion by directly including a population of *superparticles* with initial velocity distribution chosen to approximate the desired velocity distribution function $f(\mathbf{v})$.

The approach is unusual in that it mixes the Lagrangian (where individual particles are tracked directly) and Eulerian (where the flow of particles is tracked on a fixed grid) reference frames. Connecting the two frames involves two steps, *interpolation* and *projection*. In the projection step, each particle is assumed to have some finite-sized shape defined by a shape function $S(\mathbf{r})$, which is used to project the particle charge ρ and current \mathbf{J} from the particle location \mathbf{r}_i onto some set of nearby grid locations. In the interpolation step, the particle is moved under the influence of the fields \mathbf{E} and \mathbf{B} , where the fields must be interpolated at the location of the particle. The basic process is illustrated in Figure 1.11.

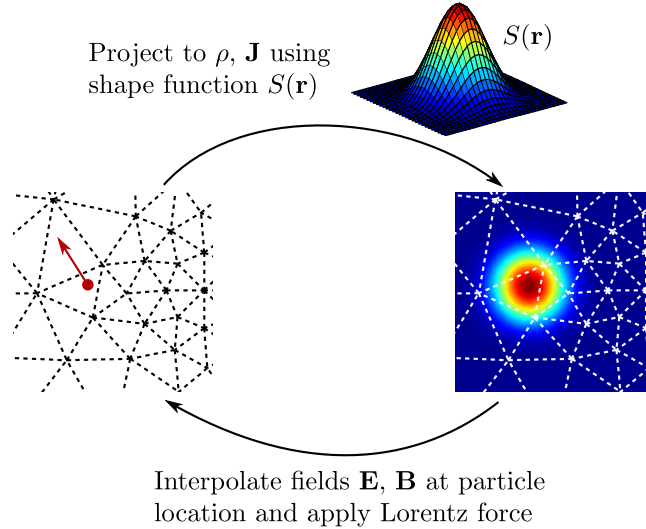


Figure 1.11: Illustration of the PIC concept.

Particle-in-cell methods are extremely flexible, working even when the plasma approximation is not strictly applicable. However, the technique has high levels of artificial noise which can only be mitigated by using a very large number of particles,

or by using artificially quiet *cold start* particle distributions. We describe our DG-PIC technique in great detail in Chapter 4.

1.10.3 Multi-moment fluid codes

Another approach to modeling plasmas is to take averages of the Vlasov equation, or so-called *moments*, as described in detail in Section 2.1.5. Each higher moment of the Vlasov equation captures more detail about the distribution: the zeroth moment captures the average charge density, the first moment captures the average current, the second moment captures the temperature, and so on. The technique leads to very quiet, robust descriptions of plasmas near equilibrium but fails in all other cases. In addition, the number of terms in the equation increases exponentially with the number of moments, practically limiting the method to four moments or less [Chevalier et al., 2010].

1.10.4 Direct Vlasov simulation

The full Vlasov equation (Equation 2.12) is a six-dimensional description of the time evolution of a particle distribution. Direct techniques, which discretize and track the entire six-dimensional space, can be used but are extremely limited due to the very large number of unknowns involved. For instance, using a naive discretization with 32 cells per velocity component and 100 cells per spatial component, over 30 billion unknowns are required just to track the distribution function. At present, simulations of this magnitude are nearly intractable, so direct methods are usually restricted to small, reduced-dimension problems. 1d3v simulations (1 spatial dimension, 3 velocity dimensions) are more feasible, but attaining competitive speeds (relative to PIC) requires sophisticated adaptive techniques [Mehrenberger et al., 2006; Sonnendrücker et al., 2004, 2005]. We do not investigate this approach in this dissertation, although we acknowledge that the technique is promising.

1.11 History of numerical modeling

The history of numerical modeling of electromagnetic wave propagation is interesting and varied. New developments in numerical modeling frequently paralleled advances

in computer technology. One of the first applications of early computers was the computation of artillery firing tables, which listed the range and trajectories of a projectile fired from a large gun or cannon under varying crosswinds or downwinds. This problem requires the solution of an ordinary differential equation (ODE) for a wide variety of parameters and initial conditions. Interestingly, this problem has a direct electromagnetic analogy: the trajectory followed by a light ray in a smoothly-varying medium is governed by a very similar ODE.

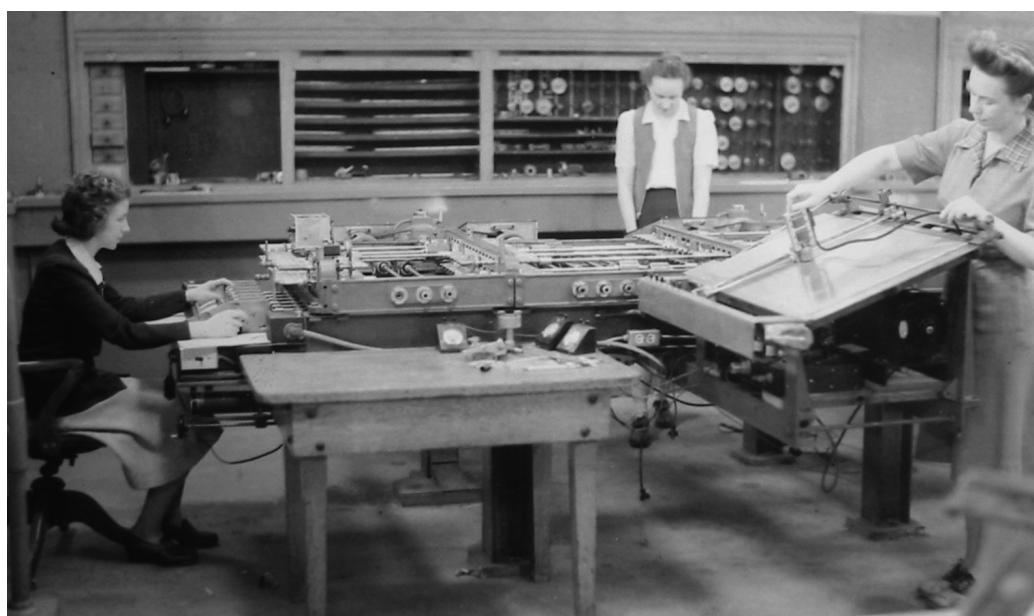


Figure 1.12: Kay McNulty, Alyse Snyder, and Sis Stump operating a differential analyzer at the University of Pennsylvania. *Image in the public domain in the United States because it is a work of the United States Federal Government under the terms of Title 17, Chapter 1, Section 105 of the US Code.*

Prior to the development of the general-purpose computer, such firing tables were computed by hand, an extremely laborious task requiring many dozens of skilled workers. The first major advance was the development of machines that could solve such ODEs. One such early machine, Vannevar Bush's differential analyzer [[Bush and Hazen, 1931](#)], could compute a trajectory in a small fraction of the time of a human specialist. The machine was essentially a collection of mechanical analogues of mathematical operations such as integration, multiplication, and addition, each one of them an ingenious design all its own. The strength of the device, however, was that

each of these fundamental blocks could be physically connected together to program a specific ODE for solution. The only essential limit to the complexity of an ODE that could be solved was the number of blocks available. However, the device was difficult to program. Intermediate blocks had to be carefully tuned to balance accuracy against the physical limits of the machine. For instance, very divergent trajectories or large swings could physically clamp parts of the machine against internal mechanical stops. Furthermore, the machine was highly specialized only for one task, which motivated the development of more general-purpose electronic computers, such as the ENIAC, an early programmable computer that used vacuum tubes for switching.

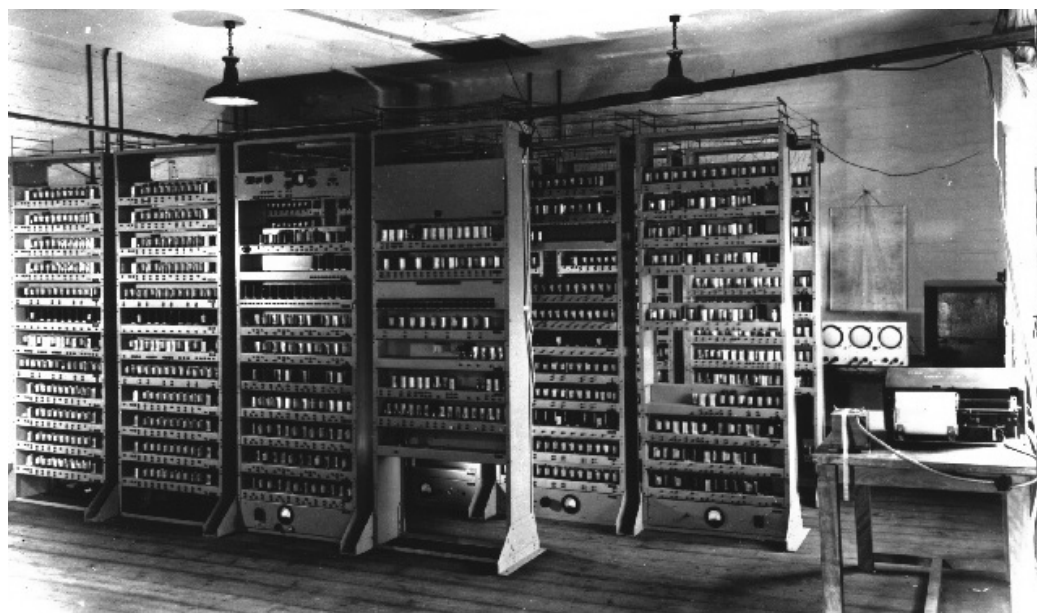


Figure 1.13: Photograph of the EDSAC I in June 1948. *Copyright Computer Laboratory, University of Cambridge. Reproduced by permission.*

While the ENIAC was also primarily designed to compute firing tables [[Kempf, 1961](#), Chapter 2], it was a fully general-purpose computing platform, programmable by a combination of punch cards and physical rewiring. One thing the ENIAC sorely lacked, however, was a large programmable internal memory, which motivated the development of a stored-program computer, the EDVAC. The innovations and design of the EDVAC were summarized by John von Neumann in the document *First Draft of a report on the EDVAC* [[von Neumann, 1945](#)]. His clear, logical description of the

design of general-purpose computing machines had enormous influence on the then-new field of computation, inspiring British researchers to construct their own model based on his document. Their resulting machine, the EDSAC (Figure 1.13), was the first general-purpose stored-program computer, that is, it could store programs in memory instead of relying on sequential punch cards. It was on this computing platform that Jenifer Haselgrove, at the time a young graduate student at Cambridge University, carried out one of the first electromagnetic simulations ever performed on a modern-style computing platform [Hartley, 1999]. Specifically, she designed a program to compute the paths traversed by monochromatic electromagnetic rays in a magnetized plasma [Haselgrove, 1955] by framing the problem in Hamiltonian theory, exactly the same framework used to compute ballistic trajectories.

The potential of the computer was, of course, recognized by experts in a number of fields outside ballistics, leading to the rapid development of many different numerical techniques along seemingly-unrelated lines.

Much of the earliest work in the numerical solution of partial differential equations (PDEs) focused on the finite difference method, which works by approximating the continuous spatial (and possibly temporal) derivatives in the PDE with finite differences. Proving stability, accuracy, and even existence of the finite difference solution dominated the early literature. It was during this time that Lax, Courant, and Wendroff did the pioneering early work on methods that now bear their names [Lax and Keller, 1951; Lax and Richtmyer, 1956, republished 2005; Lax, 1958; Lax and Wendroff, 1964, republished 2005, 1960; Courant et al., 1952; Weinberger, 1959].

In 1966, drawing inspiration from the already large body of extant literature on finite difference methods, K. Yee wrote his seminal paper describing a finite difference method for the solution of Maxwell's equations on arbitrary three-dimensional domains [Yee, 1966]. Interestingly, Yee's method for electromagnetic problems was largely abandoned for many years in favor of variational and moment methods because of their particular strength in solving scattering and boundary value problems on the limited computing resources of the time.

At the same time, Russian researchers such as *Tikhonov and Samarskii* [1960] began their pioneering work on conservative finite difference schemes, which would

later develop into what are now known as *finite volume methods*. Finite volume methods are extremely flexible and powerful, particularly for problems in nonlinear fluid flow and thus completely dominated the fluid dynamics field until only very recently.

The first precursors to the finite element method (as it is known in its modern form) date back to as early as 1941, when [Hrennikoff \[1941\]](#) modeled the behavior of a large mechanical system by dividing the structure into a mesh of finite deformable rods, which, when interconnected, approximated the behavior of the whole.

Much of the progress in the field of numerical simulation has been in combining seemingly unrelated methods, or by showing that existing methods can be generalized to solve larger classes of problems. For instance, the finite element method was originally derived by [Hrennikoff \[1941\]](#) in a very ad-hoc way, by quite literally modeling a mechanical system as a mesh of small, interconnected springs. Later authors extended the method and put it on a more sound theoretical foundation [[Strang and Fix, 2008](#)]. The Runge-Kutta method [[Runge, 1912](#), p. 133] dates back to before 1912 and was originally designed to speed up graphical calculations of the solution of initial-value problems. The technique was only placed in a more general framework many years later, as its usefulness in computation became apparent. Today the Runge-Kutta method is an entire family of highly flexible time-stepping techniques with variable order, full splitting, and embedding for use in adaptive time-stepping. The Fourier series, developed by Fourier in 1807 to solve the heat equation subject to boundary conditions, is now known to be one example of a broader class of *spectral methods*, which work by projecting the unknown infinite-dimensional solution of a PDE onto a finite-dimensional basis, effectively solving the problem in a truncated *inner product space*. The discontinuous Galerkin method, which is the focus of this dissertation, is essentially a hybrid technique combining features from both the finite volume method and the finite element method to yield a scheme that is both highly local and highly accurate.

Development of numerical methods has also been highly influenced by changes in computing technology. In recent years, direct, global methods like the finite difference method and DG method have seen a resurgence of interest for solving problems in

electromagnetics. This resurgence is largely due to the ready availability of cheap RAM, fast processors, and ubiquitous fast interconnects. Today it is perfectly feasible to directly simulate very large-scale three dimensional electromagnetics problems.

1.12 Commentary

A major focus of this dissertation is the application of the DG method to modeling VLF wave propagation in plasmas. Our primary goal is to ultimately reduce the constraints or limitations placed on the modeler. Considered in this context, the DG method has a number of advantages over other global techniques like the finite difference method and the finite volume method, namely its geometric flexibility, strong handling of material discontinuities, and arbitrary order of accuracy.

Since full, global simulations of electromagnetic wave propagation over the entire near-Earth space environment are still intractable, we limit our focus specifically to linear and nonlinear electromagnetic wave propagation in strongly inhomogeneous media. We prefer to make as few as possible a priori assumptions about the problem geometry or the spatial distribution of the plasma's physical parameters. We do not make approximations that impose geometric constraints or constraints on smoothness of the solution or background medium. The only significant constraint our method requires is a sampling constraint, that is, the phenomena of interest must be sufficiently sampled in both space and time.

In short, we stress generality over efficiency, although we do note that the DG method can easily outperform other, simpler time domain methods such as the finite difference time domain (FDTD) method, due to its high accuracy and geometric flexibility. In addition, the high accuracy of the method means that it can be used to validate more specialized approximate techniques (such as those using linearizations), offering an independent evaluation of their range of validity and manner of breakdown.

Chapter 2

Science background

2.1 Basic plasma theory

A *plasma* refers to a partially or completely ionized gas that meets a few basic criteria:

- A plasma is quasi-neutral, that is, it is composed of negatively and positively charged particles in roughly equal quantities. Local deviations from quasi-neutrality are allowed, but these must be small relative to the total number density in the plasma.
- The characteristic length scales must be bigger than a Debye length, that is, the length scales considered must be large enough so that charge screening can occur. Behavior on sizes smaller than a Debye length is dominated by statistical fluctuations, and the plasma description cannot apply on these scales.
- The charged particles must be unbound, that is, not bound to any particular nucleus and thus subject only to external forces and the Lorentz force.

An overriding theme here is that plasmas behave essentially as fluids. The number of individual particles is so large that the behavior of the system as a whole can be described quite independently of the behavior of its constituent particles. Nevertheless, a major component of this dissertation is devoted to the particle-in-cell (PIC) method, which models the behavior of a plasma by tracking the motion of individual quasi-particles, so we continue with a brief overview of single-particle motion in plasmas.

2.1.1 Definitions and single-particle motion

All plasma behavior ultimately derives from the Lorentz force law, which gives the force exerted on a particle of charge q and moving at velocity \mathbf{v} by electric and magnetic fields \mathbf{E} and \mathbf{B} , respectively:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (2.1)$$

A plasma consists of a quasi-neutral collection of both positive and negative charges. If the electrons and ions in a slab of plasma are displaced from each other, an electric field is set up which (according to the Lorentz force) will act to push the particles back towards equilibrium. In the absence of any losses or collisions, the system continues to oscillate forever about its equilibrium position at a frequency called the *plasma frequency*:

$$\omega_p = \sqrt{\frac{Nq^2}{m\epsilon_0}}, \quad (2.2)$$

where m is the particle mass, ϵ_0 is the permittivity of free space, and N is the number density in particles per cubic meter.

We primarily consider cases with a static background magnetic field \mathbf{B}_0 . In the absence of any electric field, the Lorentz force only acts at right angles to the particle velocity and background magnetic field, meaning that a particle with any initial velocity gyrates around the static magnetic field lines at the *gyrofrequency* ω_c :

$$\omega_c = \frac{qB_0}{m}, \quad (2.3)$$

where B_0 is the background magnetic field magnitude. We use the convention that the charge q carries a sign. A negative gyrofrequency (for electrons) means the rotation is in the right-hand sense with respect to the magnetic field, while a positive gyrofrequency (for ions) means the rotation is in the left-hand sense.

It is occasionally convenient to define an oriented *gyrofrequency vector*:

$$\boldsymbol{\omega}_c = \frac{q\mathbf{B}_0}{m} \quad (2.4)$$

Here \mathbf{B}_0 is the vector magnetic field, and thus $\boldsymbol{\omega}_c$ is also a vector with magnitude equal to the gyrofrequency and pointing in the direction of the axis of rotation. As above, the sign of q denotes the sense of the rotation with respect to the magnetic field.

For wave-particle interactions, we consider the response of a particle to time-harmonic electromagnetic fields, that is:

$$\begin{aligned} \mathbf{E}(t) &= \mathbf{E}e^{-j\omega t} \\ \mathbf{B}(t) &= \mathbf{B}e^{-j\omega t} \end{aligned}$$

Under such forced motion, individual particles can execute complicated trajectories, as shown in Figure 2.1 for circularly polarized harmonic forcing.

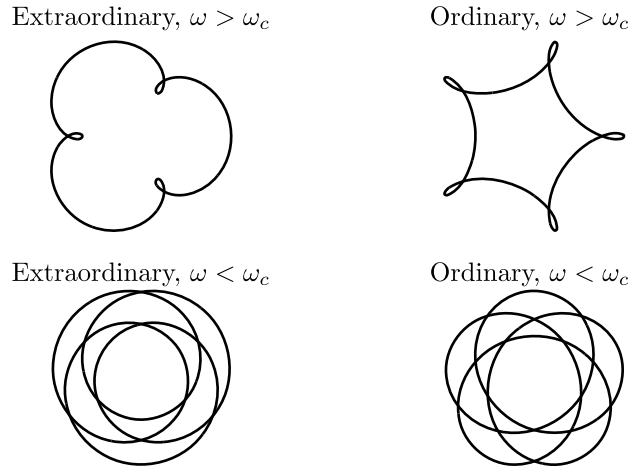


Figure 2.1: Individual particle motion under forcing from a circularly polarized electric field, with static background magnetic field oriented out of the plane of the plot. Four cases are shown, for wave frequencies above the particle gyrofrequency in the top two panels and below the gyrofrequency for the bottom two. The left and right sets of panels are for different polarizations of the wave, either leading to driven motion rotating in the same sense as the particle gyromotion or in the opposite sense.

This forced motion constitutes an additional current, which is not necessarily

in the same direction or of the same phase as the input wave. Since the plasma is composed of many such particles, these additional currents in turn modify the properties of the wave propagating through the medium. Supposing that we have an ensemble of N particles indexed by i , the electromagnetic behavior of a plasma is (in principle) completely described by a relatively simple set of equations:

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E} \quad (2.5a)$$

$$\mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{B} - \mu_0 \mathbf{J} \quad (2.5b)$$

$$\mathbf{J} = \sum_{i=1}^N \mathbf{v}_i q_i \quad (2.5c)$$

$$\rho = \sum_{i=1}^N q_i \quad (2.5d)$$

subject to the additional constraints:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \quad (2.6a)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (2.6b)$$

While this discrete-particle description in principle captures all plasma phenomena, in practice it is useless except in very simple cases, due to the extremely large number of particles in natural plasmas of any interesting size – for instance, charged particle number densities on the order of $N = 10^6 \text{ m}^{-3}$ or larger are commonplace in the Earth’s ionosphere. Since typical scale sizes of interest for radio propagation in the ionosphere are on the order of hundreds of km^3 , direct simulation of every particle in the system is currently intractable. As such, it is common to treat the plasma as a continuous fluid for analysis and numerical modeling. We will return to the single-particle approach in Chapter 4 when discussing the PIC approach, but we will do so in the sense of approximating the fluid description.

2.1.2 Distribution function

The problems in modeling such a large number of particles can be remedied by recognizing that particles are interchangeable. In other words, a particle traveling at a given velocity \mathbf{v} and at a position \mathbf{r} is indistinguishable from another particle at the same velocity and position. Likewise, two particles at different positions and velocities could be interchanged at will with absolutely no influence on the results.

Thus, it is natural to consider the entire population of particles in a statistical sense. Instead of a discrete population of particles evolving self-consistently as time progresses, we consider a distribution of particles evolving under the same set of physical laws. For a given particle, its state is entirely determined by its velocity and position, so the plasma distribution function is defined over a six-dimensional space (three velocity and three position coordinates) termed *phase space*.

The phase space distribution function f is defined such that f is related to the infinitesimal number density of particles dN in an infinitesimal phase space volume $d\mathbf{r} d\mathbf{v}$ via:

$$dN = f(\mathbf{r}, \mathbf{v}, t) d\mathbf{r} d\mathbf{v}$$

The distribution function, then, tells us the density of particles at a given position \mathbf{r} moving at a velocity \mathbf{v} . The total number of particles N is found by integrating:

$$N(t) = \iint f(\mathbf{r}, \mathbf{v}, t) d\mathbf{r} d\mathbf{v}$$

And the number density is then:

$$N(\mathbf{r}, t) = \int f(\mathbf{r}, \mathbf{v}, t) d\mathbf{v} \tag{2.7}$$

We can continue this process, multiplying the distribution function by higher and higher powers of the velocity and integrating to find the *moments* of the distribution.

For instance, the first moment is proportional to the average bulk velocity \mathbf{u} :

$$u_i(\mathbf{r}, t) = \frac{1}{N} \int v_i f(\mathbf{r}, \mathbf{v}, t) d\mathbf{v} \quad (2.8)$$

The second moment is the pressure tensor:

$$P_{ij}(\mathbf{r}, t) = \int m(v_i - u_i)(v_j - u_j) f(\mathbf{r}, \mathbf{v}, t) d\mathbf{v} \quad (2.9)$$

The third moment gives the heat flux tensor:

$$Q_{ijk}(\mathbf{r}, t) = \int m(v_i - u_i)(v_j - u_j)(v_k - u_k) f(\mathbf{r}, \mathbf{v}, t) d\mathbf{v} \quad (2.10)$$

Each higher moment captures additional information about the distribution in some average sense. However, since the number of terms in the tensors grow exponentially, it is extremely cumbersome to consider moments higher than the third or fourth.

By coupling a force term into the continuity equation for phase space density, we can write an equation for the time evolution of the distribution function. If this force is the Lorentz force, this equation is called the Vlasov equation, which we discuss next.

2.1.3 The Vlasov approximation

Consider a two-dimensional system, with only a scalar position x and scalar velocity v . A particle begins at state (x, v) subject to an external force F . An infinitesimal time dt later, the particle will be at position $(x + v dt, v + (F/m) dt)$, as shown in Figure 2.2.

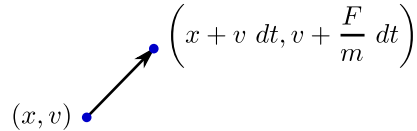


Figure 2.2: Trajectory of a single particle in phase space before and after an infinitesimal timestep dt .

It is clear that we can consider the vector $(v, F/m)$ as the direction of some

generalized flow of particles through phase space. Multiplying this vector by the phase space density f , we immediately recognize $\phi = (fv, fF/m)$ as a pointwise flux, representing the flow of a density f through phase space. Assuming that particles are conserved, we can directly write a conservation law in these coordinates (x, v) :

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial x}(fv) + \frac{\partial}{\partial v}\left(f\frac{F}{m}\right) = 0$$

The extension to six dimensions follows straightforwardly. In vector form:

$$\frac{\partial f}{\partial t} + \nabla \cdot \phi = 0 \quad (2.11)$$

where

$$\begin{aligned} \nabla &= \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial y}, \frac{\partial}{\partial v_x}, \frac{\partial}{\partial v_y}, \frac{\partial}{\partial v_z} \right) \\ \phi &= \left(fv_x, fv_y, fv_z, f\frac{F_x}{m}, f\frac{F_y}{m}, f\frac{F_z}{m} \right) \end{aligned}$$

In this form, it is clear that this is simply a statement of conservation of phase space density. Substituting the Lorentz force for the term \mathbf{F} , we have the Vlasov equation in conservative form:

$$\frac{\partial f}{\partial t} + \nabla_r \cdot (f\mathbf{v}) + \nabla_v \cdot \left[f\frac{q}{m}(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \right] = 0 \quad (2.12)$$

∇_r and ∇_v denote the gradients with respect to space and velocity, respectively. The Vlasov equation is commonly rewritten in non-conservative form as follows. The divergence of the Lorentz force with respect to velocity is always zero, $\nabla_v \cdot \mathbf{F} = 0$, which is easily verified. Assuming the particle mass m is constant (which holds true for non-relativistic distributions), the divergence of the acceleration $\mathbf{a} = \mathbf{F}/m$ is therefore also zero, that is, $\nabla_v \cdot \mathbf{a} = 0$. Second, each term of $\nabla_r \cdot \mathbf{v}$ is identically zero, since \mathbf{v} and \mathbf{r} are independent variables. Therefore,

$$\nabla_r \cdot \mathbf{v} + \nabla_v \cdot \mathbf{a} = 0$$

Another way of stating this result is that the distribution is incompressible in phase space. We note that this condition only holds in the absence of any loss mechanisms such as particle sinks, sources, or collisional losses. Expanding the product terms and simplifying, we obtain the Vlasov equation in non-conservative form:

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_r f + \frac{q}{m} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \nabla_v f = 0 \quad (2.13)$$

The Vlasov equation is rarely used directly. For some simplified cases (e.g., beam instabilities), it can be used directly, but for analysis, the equation is usually first simplified by linearization or by taking successive fluid moments.

2.1.4 Linearization of the Vlasov equation

Linearizing the Vlasov equation is a common first step in determining the small-signal behavior of waves in a hot plasma. We assume that all quantities of interest consist of an equilibrium value plus a small first-order perturbation:

$$f(\mathbf{v}) = f_0(\mathbf{v}) + f_p(\mathbf{v})$$

$$\mathbf{E} = 0 + \mathbf{E}_w$$

$$\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_w$$

The first-order perturbation on the distribution function f_p is due to a wave propagating with electric and magnetic fields \mathbf{E}_w and \mathbf{B}_w . \mathbf{B}_0 is the static background magnetic field. We first note that the zeroth-order terms must themselves satisfy the Vlasov equation in the absence of any perturbation [*Inan and Golkowski, 2010*, p. 241], that is:

$$\frac{\partial f_0}{\partial t} + \mathbf{v} \cdot \nabla_r f_0 + \frac{q}{m} (\mathbf{E}_0 + \mathbf{v} \times \mathbf{B}_0) \cdot \nabla_v f_0 = 0$$

Substituting these equations into Equation 2.13 and dropping all products of small terms, we have:

$$\frac{\partial f_p}{\partial t} + \mathbf{v} \cdot \nabla_r f_p + \frac{q}{m} (\mathbf{v} \times \mathbf{B}_0) \cdot \nabla_v f_p + \frac{q}{m} (\mathbf{E}_w + \mathbf{v} \times \mathbf{B}_w) \cdot \nabla_v f_0 = 0 \quad (2.14)$$

The linearized Vlasov equation is a powerful tool for determining the small-signal growth and damping rates of propagating modes in a plasma.

2.1.5 Moments of the Vlasov equation

Similar to the moments of the distribution function taken in Section 2.1.2, we can also take moments of the Vlasov equation to find equations describing the time evolution of the various moments of the distribution (number density, pressure, heat flux, etc.). The derivations herein closely follow those in *Inan and Golkowski* [2010, Chapter 4].

The zeroth moment is found by integrating each term of Equation 2.12:

$$\frac{\partial}{\partial t} \int f d\mathbf{v} + \int \nabla_r \cdot (f\mathbf{v}) d\mathbf{v} + \frac{1}{m} \int \nabla_v \cdot (f\mathbf{F}) d\mathbf{v} = 0$$

The first term is equal to the time rate of change of the number density:

$$\frac{\partial}{\partial t} \int f d\mathbf{v} = \frac{\partial N}{\partial t}$$

For the second term, we interchange the integral and derivative as they apply to different coordinates:

$$\nabla_r \cdot \int f\mathbf{v} d\mathbf{v} = \nabla \cdot (N\mathbf{u}),$$

where \mathbf{u} is the average bulk velocity. By physical argument, the third term is zero, which we can see by rewriting the integral as a surface integral:

$$\int \nabla_v \cdot (f\mathbf{F}) d\mathbf{v} = \oint f\mathbf{F} \cdot d\mathbf{s}_v$$

Any realistic density f will vanish quickly enough as $|\mathbf{v}| \rightarrow \infty$ so the magnitude of

this term must necessarily go to zero. Thus,

$$\frac{\partial N}{\partial t} + \nabla \cdot (N\mathbf{u}) = 0, \quad (2.15)$$

which is a statement of conservation of charge. Similarly, the first moment is:

$$\frac{\partial}{\partial t} \int v_i f \, d\mathbf{v} + \int v_i \nabla_r \cdot (f\mathbf{v}) \, d\mathbf{v} + \frac{1}{m} \int v_i \nabla_v \cdot (f\mathbf{F}) \, d\mathbf{v} = 0$$

The first term is:

$$\frac{\partial}{\partial t} \int v_i f \, d\mathbf{v} = \frac{\partial}{\partial t} (N u_i)$$

The second term:

$$\begin{aligned} \int v_i \nabla_r \cdot (f\mathbf{v}) \, d\mathbf{v} &= \frac{\partial}{\partial x_j} \int v_i v_j f \, d\mathbf{v} \\ &= \frac{\partial}{\partial x_j} \int [(v_j - u_j)(v_i - u_i) + (v_i u_j + v_j u_i - u_i u_j)] f \, d\mathbf{v} \\ &= \frac{\partial}{\partial x_j} \left(\frac{1}{m} P_{ij} + N u_i u_j \right) \end{aligned}$$

Thus,

$$\int \mathbf{v} \nabla_r \cdot (f\mathbf{v}) \, d\mathbf{v} = \frac{1}{m} \nabla_r \cdot \mathbf{P} + \nabla_r \cdot (N\mathbf{u}\mathbf{u}),$$

where the pressure tensor \mathbf{P} is defined in Equation 2.9 and $\mathbf{u}\mathbf{u}$ is the dyadic product defined in Section 1.3. The third term is more complicated. Evaluating the integral by parts,

$$\frac{1}{m} \int v_i \nabla_v \cdot (f\mathbf{F}) \, d\mathbf{v} = \frac{1}{m} \oint v_i f \mathbf{F} \cdot d\mathbf{s}_v - \frac{1}{m} \int (f\mathbf{F}) \cdot (\nabla_v v_i) \, d\mathbf{v}$$

Once again, we argue on physical grounds that the density f decays sufficiently fast as $|\mathbf{v}| \rightarrow \infty$ that the surface integral term goes to zero. For the second term, we have

two cases:

$$-\frac{1}{m} \int (f\mathbf{F}) \cdot (\nabla_v v_i) d\mathbf{v} = -\frac{1}{m} \int f F_j \frac{\partial v_i}{\partial v_j} d\mathbf{v}$$

When $i \neq j$, the integral is identically zero since $\partial v_i / \partial v_j = 0$.

When $i = j$, $\partial v_i / \partial v_j = 1$, so we have:

$$-\frac{1}{m} \int f\mathbf{F} d\mathbf{v} = -\frac{Nq}{m}(\mathbf{E} + \mathbf{u} \times \mathbf{B}),$$

where we have substituted the Lorentz force for \mathbf{F} . We now have:

$$\frac{\partial}{\partial t}(N\mathbf{u}) + \frac{1}{m} \nabla_r \cdot \mathbf{P} + \nabla_r \cdot (N\mathbf{u}\mathbf{u}) - N\frac{q}{m}(\mathbf{E} + \mathbf{u} \times \mathbf{B}) = 0 \quad (2.16)$$

We can further simplify Equation 2.16 to put the temporal and spatial derivatives in a more familiar form, yielding the first moment of the Vlasov equation:

$$N \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) + \frac{1}{m} \nabla_r \cdot \mathbf{P} - N\frac{q}{m}(\mathbf{E} + \mathbf{u} \times \mathbf{B}) = 0 \quad (2.17)$$

We can continue to higher and higher moments, but once again, the number of terms grows exponentially so it is rarely practical to continue past the third or fourth moment. An additional problem is closure, that is, each moment introduces an unknown quantity. For instance, Equation 2.17 contains a pressure term \mathbf{P} , but we have no equation telling us how this quantity evolves. The second moment would provide a time evolution equation for \mathbf{P} , but it introduces a new unknown, the heat flux \mathbf{Q} . At some point, we must appeal to some physical assumption in order to close the system of equations, e.g., by assuming that the temperature is constant. One important example of a simple closure relationship is the *cold plasma approximation*, discussed in the next section.

2.1.6 Cold plasma approximation

We begin with Equation 2.17, which has an unknown pressure tensor \mathbf{P} . In the cold plasma approximation, we assume the temperature (and therefore pressure) are zero

and this term is simply dropped. Further, for small signal analysis, it is common to drop the small convective derivative term $\mathbf{u} \cdot \nabla \mathbf{u}$. Identifying $qN\mathbf{u}$ as the current \mathbf{J} , we have:

$$\frac{d\mathbf{J}}{dt} - \frac{q^2 N}{m} \mathbf{E} - \frac{q}{m} \mathbf{J} \times \mathbf{B} = 0$$

If we further assume that the background magnetic field \mathbf{B}_0 is much larger than any wave magnetic field \mathbf{B}_w , we can drop the cross term $\mathbf{J} \times \mathbf{B}_w$. Finally, using Equations 2.2 and 2.4, we rewrite the equation in a more compact form:

$$\frac{d\mathbf{J}}{dt} = \epsilon_0 \omega_p^2 \mathbf{E} - \boldsymbol{\omega}_c \times \mathbf{J}$$

Although we have not discussed collisions, it is also common to introduce an effective collision frequency representing macroscopic losses:

$$\frac{d\mathbf{J}}{dt} + \nu \mathbf{J} = \epsilon_0 \omega_p^2 \mathbf{E} - \boldsymbol{\omega}_c \times \mathbf{J} \quad (2.18)$$

Under this set of approximations, a cold plasma is completely characterized by a first-order ODE relating the cold plasma current \mathbf{J} to the input \mathbf{E} and the static background magnetic field \mathbf{B}_0 . While plasmas are not cold, Equation 2.18 is a reasonable approximation of Equation 2.17 in certain regimes, in particular in the Earth's ionosphere.

2.1.7 The cold plasma conductivity tensor

We begin with Equation 2.18. For notational convenience, we denote the complex frequency $s = -j\omega$. Making the substitution $d/dt \rightarrow s$, we have a matrix equation relating the cold plasma current \mathbf{J} and the electric field \mathbf{E} :

$$\begin{bmatrix} s + \nu & -\omega_{cz} & \omega_{cy} \\ \omega_{cz} & s + \nu & -\omega_{cx} \\ -\omega_{cy} & \omega_{cx} & s + \nu \end{bmatrix} \mathbf{J} = \epsilon_0 \omega_p^2 \mathbf{E}$$

Symbolically inverting this matrix yields a conductivity tensor relating \mathbf{J} and \mathbf{E} via $\mathbf{J} = \vec{\sigma} \mathbf{E}$, where each entry is a ratio of polynomials in s :

$$\vec{\sigma}(s) = \frac{\epsilon_0 \omega_p^2}{d(s)} \begin{bmatrix} \sigma_{xx}(s) & \sigma_{xy}(s) & \sigma_{xz}(s) \\ \sigma_{yx}(s) & \sigma_{yy}(s) & \sigma_{yz}(s) \\ \sigma_{zx}(s) & \sigma_{zy}(s) & \sigma_{zz}(s) \end{bmatrix} \quad (2.19)$$

Where:

$$d(s) = (s + \nu)(s^2 + 2s\nu + \nu^2 + |\omega_c|^2)$$

and:

$$\begin{aligned} \sigma_{xx}(s) &= s^2 + 2\nu s + \nu^2 + \omega_{cx}^2 \\ \sigma_{xy}(s) &= \omega_{cz}s + \omega_{cx}\omega_{cy} + \nu\omega_{cz} \\ \sigma_{xz}(s) &= -\omega_{cy}s - \nu\omega_{cy} + \omega_{cx}\omega_{cz} \\ \sigma_{yx}(s) &= -\omega_{cz}s + \omega_{cx}\omega_{cy} - \nu\omega_{cz} \\ \sigma_{yy}(s) &= s^2 + 2\nu s + \nu^2 + \omega_{cy}^2 \\ \sigma_{yz}(s) &= \omega_{cx}s + \nu\omega_{cx} + \omega_{cy}\omega_{cz} \\ \sigma_{zx}(s) &= \omega_{cy}s + \nu\omega_{cy} + \omega_{cx}\omega_{cz} \\ \sigma_{zy}(s) &= -\omega_{cx}s - \nu\omega_{cx} + \omega_{cy}\omega_{cz} \\ \sigma_{zz}(s) &= s^2 + 2\nu s + \nu^2 + \omega_{cz}^2 \end{aligned}$$

If multiple plasma species are present, the currents simply sum by linearity:

$$\mathbf{J}(s) = (\vec{\sigma}_1(s) + \vec{\sigma}_2(s) + \dots) \mathbf{E}(s),$$

where $\vec{\sigma}_j(s)$ is the conductivity tensor for the j^{th} plasma species.

2.2 Waves in plasmas

The steady-state analysis of waves in plasmas leads to a *dispersion relation*, which relates the wave vector \mathbf{k} to the wave frequency ω . We begin by transforming the field quantities in Maxwell's equations into the frequency domain. Assuming a plane wave solution of the form:

$$\mathbf{E}(t) = \mathbf{E}e^{j(\mathbf{k}\cdot\mathbf{r}-\omega t)}$$

$$\mathbf{H}(t) = \mathbf{H}e^{j(\mathbf{k}\cdot\mathbf{r}-\omega t)}$$

The differential operators transform as:

$$\frac{\partial}{\partial t} \rightarrow -j\omega \quad (2.20a)$$

$$\nabla \rightarrow j\mathbf{k} \quad (2.20b)$$

When considering losses or gains, the frequency ω (or \mathbf{k}) are complex, that is, $\omega = \omega_r + j\omega_i$. The imaginary component ω_i thus represents a gain (positive) or a loss (negative).

We begin with waves in cold plasmas and follow with a discussion of Landau damping and whistler mode cyclotron instabilities. The cold plasma dispersion relation can be solved easily. In a hot plasma, by contrast, the concept of a mode only makes sense for smooth distribution functions considered under long time behavior, but it is nonetheless a useful technique for understanding anisotropy-driven wave instabilities in plasmas.

2.2.1 The cold plasma dispersion relation

We begin by writing Maxwell's equations for a cold plasma in the frequency domain, where we have transformed the differential operators as in Equation 2.20:

$$\mathbf{k} \times \mathbf{E} = \omega\mu_0\mathbf{H}$$

$$\mathbf{k} \times \mathbf{H} = -\omega \left(\epsilon_0 \mathbf{I} - \frac{\vec{\sigma}_1(\omega)}{j\omega} - \frac{\vec{\sigma}_2(\omega)}{j\omega} - \dots \right) \mathbf{E}$$

Identifying the term in parentheses as an effective permittivity $\vec{\epsilon}_{\text{eff}}(\omega)$, we have:

$$\mathbf{k} \times \mathbf{E} = \omega \mu_0 \mathbf{H} \quad (2.21a)$$

$$\mathbf{k} \times \mathbf{H} = -\omega \vec{\epsilon}_{\text{eff}}(\omega) \mathbf{E} \quad (2.21b)$$

For convenience, we fix the propagation angle and solve for the wave vector magnitude k . We define the matrix \mathbf{K} , where we have rewritten the vector \mathbf{k} in spherical coordinates with ϕ being the angle from the $\hat{\mathbf{z}}$ axis and θ being the azimuthal angle:

$$\mathbf{K} = k \hat{\mathbf{K}} = k \begin{bmatrix} 0 & -\cos \phi & \sin \theta \sin \phi \\ \cos \phi & 0 & -\cos \theta \sin \phi \\ -\sin \theta \sin \phi & \cos \theta \sin \phi & 0 \end{bmatrix}$$

The Maxwell system in block matrix form is then:

$$\begin{bmatrix} -\omega \vec{\epsilon}_{\text{eff}}(\omega) & 0 \\ 0 & \omega \mu_0 \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{E} \\ \mathbf{H} \end{bmatrix} = k \begin{bmatrix} 0 & \hat{\mathbf{K}} \\ \hat{\mathbf{K}} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{E} \\ \mathbf{H} \end{bmatrix}$$

Numerically solving the resulting generalized eigenvalue problem yields the solution:

$$\mathbf{A} \mathbf{v} = k \mathbf{B} \mathbf{v},$$

which gives the wave vector magnitude k as a function of the frequency ω as well as the characteristics \mathbf{E} and \mathbf{H} associated with that mode. Two of the eigenvalues k are infinite. These arise as a result of the nullspace of the curl operator, reflecting the physical fact that adding a static curl-free field to \mathbf{E} or \mathbf{H} leaves the electromagnetic wave behavior unaffected.

The equation here would apply even if the permeability μ were also a frequency-dependent tensor. However, since μ is a simple scalar in a plasma, we can also eliminate \mathbf{H} , yielding a simpler eigenvalue problem:

$$\left[\omega^2 \mu_0 \vec{\epsilon}_{\text{eff}}(\omega) + k^2 \hat{\mathbf{K}}^2 \right] \mathbf{E} = 0 \quad (2.22)$$

These eigenvalues can be found algebraically by setting the determinant of the quantity in parentheses to zero and solving for k , which for a plasma (after simplification) yields a more explicit form as a quadratic in terms of k^2 , a form popularly known as the Appleton-Hartree-Lassen equation [[Appleton, 1928](#); [Hartree, 1931](#); [Lassen, 1927](#)]. We note that the eigenvalue approach, however, can apply to any linear medium with only small modifications.

Solving the cold plasma dispersion relation for the wavenumber k in terms of the frequency ω is relatively straightforward as shown. However, supposing we want to solve the inverse problem, that is, the dispersion relation $\omega(k)$, or ω as a function of wave vector magnitude k , the solution is less straightforward. For two or more cold plasma species there are more than four roots ω for a given wavenumber k , which means that an explicit solution cannot even be written down. However, an eigenvalue approach still exists and can be useful in some situations. This approach is discussed in Appendix [A](#).

2.2.2 The electrostatic hot plasma dispersion relation

A hot plasma supports purely electrostatic modes. In order to compute the small signal propagation velocity and damping rate, we use the linearized Vlasov equation (Equation [2.14](#)) to derive the dispersion relation for electrostatic waves. We first set k_\perp and \mathbf{B}_w to zero. Additionally, we consider an unmagnetized plasma for the time being, so $\omega_c = 0$. Noting that we are left with only componentwise derivatives and no cross-terms, we assume without loss of generality that propagation is in the $\hat{\mathbf{z}}$ direction. Thus, Equation [2.14](#) reduces to:

$$f_p = j \frac{q}{mk_z} E_z \frac{\partial f_0 / \partial v_z}{v_z - \omega / k_z} \quad (2.23)$$

From the divergence theorem and Equation [2.7](#), we have:

$$\epsilon_0 \frac{\partial E_z}{\partial z} = \rho = q \int f_p d\mathbf{v} \quad (2.24)$$

Substituting Equation 2.23 into 2.24 and making the Fourier transform substitution $\partial/\partial z \rightarrow jk_z$, we have:

$$\left(1 - \frac{q^2}{\epsilon_0 m k_z^2} \int \frac{\partial f_0 / \partial v_z}{v_z - \omega / k_z} d\mathbf{v}\right) E_z = 0$$

The trivial solution $E_z = 0$ is not useful, so the other term must be zero, which leads to:

$$1 - \frac{q^2}{\epsilon_0 m k^2} \int \frac{\partial f_0 / \partial v}{v - \omega / k} dv = 0 \quad (2.25)$$

Note that we have integrated out the v_x and v_y terms for clarity, so f_0 is now understood to be a purely one-dimensional function in terms of a scalar velocity v . An alternative form is found by integrating by parts:

$$1 - \frac{q^2}{\epsilon_0 m k^2} \int \frac{f_0(v)}{(v - \omega / k)^2} dv = 0 \quad (2.26)$$

The singularity in the denominator is not integrable, so a solution to this equation does not exist unless there are losses or gains allowed, that is, if either the wave frequency ω or the wavenumber k_z are allowed to take on complex values. Thus to carry out the integration, we must analytically continue v into the complex plane, allowing the singularity at $v = \omega / k$ to lie off the real axis, as shown in Figure 2.3.

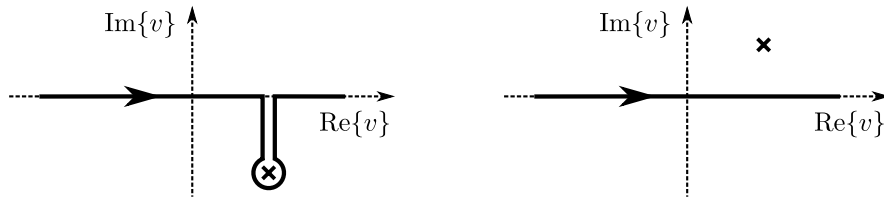


Figure 2.3: Landau integration contours in the complex v plane for the electrostatic mode in a hot plasma. The singularity $v = \omega / k$ is indicated with an x-mark. **Left:** Integration contour for damping ($\text{Im}\{\omega_i\} < 0$). **Right:** Integration contour for gain ($\text{Im}\{\omega_i\} > 0$). In both cases the wavenumber k is taken to be positive.

Carrying out the contour integration on Equation 2.25, we now have the electrostatic dispersion relation for damping in an integrable form, where v is again a real

quantity and the contour integration around the pole has contributed a factor of $2\pi j$:

$$1 - \frac{q^2}{\epsilon_0 m k^2} \left(\int_{-\infty}^{\infty} \frac{\partial f_0 / \partial v}{v - \omega/k} dv + 2\pi j \left. \frac{\partial f_0}{\partial v} \right|_{v=\omega/k} \right) = 0 \quad (2.27)$$

Equation 2.27 is valid for damping only. For gain, the contour is not deformed so the second term is zero:

$$1 - \frac{q^2}{\epsilon_0 m k^2} \int_{-\infty}^{\infty} \frac{\partial f_0 / \partial v}{v - \omega/k} dv = 0 \quad (2.28)$$

Either ω or k (or possibly both) are now complex numbers. This equation can be solved numerically using a complex root finding technique. Typically one of the frequencies, either ω or k , is chosen as some fixed, real number, and then the root finder solves for the other complex variable, i.e., to find $k(\omega)$ or $\omega(k)$.

2.2.3 The electromagnetic hot plasma dispersion relation

In principle, solving the dispersion relation for the electromagnetic modes in a linearized hot plasma is similar to the approach taken in Section 2.2.1. The primary complication is that the hot plasma equivalent permittivity tensor is non-local, that is, it depends not only on the temporal frequency ω but also the wavenumber \mathbf{k} :

$$\vec{\epsilon}_{\text{eff}}(\omega, \mathbf{k}) = \epsilon_0 \mathbf{I} - \frac{\vec{\sigma}_1(\omega, \mathbf{k})}{j\omega} - \frac{\vec{\sigma}_2(\omega, \mathbf{k})}{j\omega} - \dots$$

The actual expression for each conductivity tensor is exceedingly complicated. We reproduce it here in time-harmonic form for a one-species distribution function in cylindrical coordinates $f_0(v_{\perp}, v_{\parallel})$ without derivation. One derivation is given in [Gurnett and Bhattacharjee \[2005\]](#), pp. 367-370]:

$$\vec{\sigma} = j \frac{q^2}{m\omega_c} \sum_n \int_{-\infty}^{\infty} \int_0^{\infty} \frac{2\pi v_{\perp} dv_{\perp} dv_{\parallel}}{\alpha + n} \begin{bmatrix} a \frac{n^2 v_{\perp}}{\beta^2} J_n^2 & ja \frac{nv_{\perp}}{\beta} J_n J'_n & b \frac{nv_{\perp}}{\beta} J_n^2 \\ -ja \frac{nv_{\perp}}{\beta} J_n J'_n & av_{\perp} J'_n J'_n & -jbv_{\perp} J_n J'_n \\ a \frac{nv_{\parallel}}{\beta} J_n^2 & jav_{\parallel} J_n J'_n & bv_{\parallel} J_n^2 \end{bmatrix}$$

This expression is only valid when the background magnetic field \mathbf{B}_0 is in the $\hat{\mathbf{z}}$ direction. The subscripts \parallel and \perp denote the components parallel and perpendicular to the background magnetic field \mathbf{B}_0 , respectively. The quantities α and β are defined as:

$$\alpha = \frac{k_{\parallel}v_{\parallel} - \omega}{\omega_c}$$

$$\beta = \frac{k_{\perp}v_{\perp}}{\omega_c}$$

J_n denotes $J_n(\beta)$, the Bessel function of the first kind with argument β . J'_n denotes the derivative of the Bessel function of the first kind with respect to its argument, evaluated at β . The values a and b are defined as:

$$a = \frac{\partial f_0}{\partial v_{\perp}} + \frac{k_{\parallel}}{\omega} \left(v_{\perp} \frac{\partial f_0}{\partial v_{\parallel}} - v_{\parallel} \frac{\partial f_0}{\partial v_{\perp}} \right)$$

$$b = \frac{\partial f_0}{\partial v_{\parallel}} - \frac{n\omega_c}{\omega v_{\perp}} \left(v_{\perp} \frac{\partial f_0}{\partial v_{\parallel}} - v_{\parallel} \frac{\partial f_0}{\partial v_{\perp}} \right)$$

This expression also has a singularity of the form $k_{\parallel}v_{\parallel} - \omega$, necessitating an analytic continuation of the velocity v and an associated contour integration, as in Figure 2.3. The dispersion relation is then defined by the implicit function:

$$\det \left(\omega^2 \mu_0 \vec{\epsilon}_{\text{eff}}(\omega, k) + k^2 \hat{\mathbf{K}}^2 \right) = 0 \quad (2.29)$$

In principle, solving Equation 2.29 is not much different from solving the cold plasma dispersion relation, but is much more difficult in practice. Direct evaluation to find the modes of the system typically requires numerical techniques. A useful simplification, however, and one that can provide useful insight into cyclotron instabilities, is to assume propagation parallel to the background magnetic field. Equation 2.29 then simplifies to [*Inan and Gołkowski, 2010*, p. 242]:

$$1 - \frac{c^2 k_{\parallel}^2}{\omega^2} - \frac{q^2}{m\epsilon_0\omega^2} \int_{-\infty}^{\infty} \frac{g(v_{\parallel})}{v_{\parallel} - (\omega \pm \omega_c)/k_{\parallel}} dv_{\parallel} = 0 \quad (2.30)$$

where the sign \pm is positive for the right-hand circularly polarized mode and negative for the left-hand mode.

$$g(v_{\parallel}) = \int_0^{\infty} \left[\left(\frac{\omega}{k_{\parallel}} - v_{\parallel} \right) \frac{\partial f_0}{\partial v_{\perp}} + v_{\perp} \frac{\partial f_0}{\partial v_{\parallel}} \right] \pi v_{\perp}^2 dv_{\perp} \quad (2.31)$$

Similar to the Landau damping case, there is a non-integrable singularity. The technique to carry out the integration follows in the same manner. One final simplification is common in the literature: when growth or damping is small, only particles very near resonance with the wave contribute to the integral over v_{\parallel} , thus collapsing that integral. The imaginary component of the frequency (that is, the growth rate) can then be explicitly written [*Kennel and Petschek, 1966*]:

$$\omega_i = -\pi \frac{\omega_c}{N} \left(1 + \frac{\omega}{\omega_c} \right)^2 |v_{\text{res}}| \left(A - \frac{\omega}{|k_{\parallel} v_{\text{res}}|} \right) \int_0^{\infty} 2\pi v_{\perp} f_0(v_{\parallel} = v_{\text{res}}, v_{\perp}) dv_{\perp}, \quad (2.32)$$

where the resonant velocity is:

$$v_{\text{res}} = \frac{\omega \pm \omega_c}{k_{\parallel}} \quad (2.33)$$

and the anisotropy factor A is:

$$A = \frac{\int_0^{\infty} \frac{k_{\parallel}}{|k_{\parallel}|} \left(v_{\parallel} \frac{\partial f_0}{\partial v_{\perp}} - v_{\perp} \frac{\partial f_0}{\partial v_{\parallel}} \right) \frac{v_{\perp}^2}{v_{\parallel}} \pi dv_{\perp}}{\int_0^{\infty} 2\pi v_{\perp} f_0(v_{\parallel}, v_{\perp}) dv_{\perp}} \bigg|_{v_{\parallel}=v_{\text{res}}} \quad (2.34)$$

Sufficiently large positive anisotropy factors lead to growth (positive ω_i) of the electron cyclotron whistler mode wave. Note that we use the signed gyrofrequency defined in Equation 2.3, that is, ω_c is negative for electrons and positive for ions.

2.3 Plasma instabilities

The plasma in the Earth's magnetosphere is particularly interesting because it is frequently unstable. Because it is relatively diffuse and nearly collisionless, the plasma

in the near-Earth space environment can exist in metastable states for very long periods before releasing its energy in some manner or another. We now discuss some basic physics of instabilities from the entropy point of view and then continue with a brief discussion of some specific examples relevant in the context of this dissertation.

2.3.1 Entropy arguments

Instabilities arise in plasmas, fundamentally, because the system is not in equilibrium. From a statistical mechanics point of view, instabilities arise because the *entropy* in the system is not at a maximum, so that there is energy that could be extracted for useful work. One of the most fundamental relations in thermodynamics provides a good illustration of this concept [*Lifshitz et al.*, 1994, p. 43]:

$$dU = TdS - PdV, \quad (2.35)$$

where U is the internal energy (kinetic plus potential energy), T is the temperature, S is the entropy, P is the pressure, and V is the volume. We can see that if, for instance, the temperature, pressure, and volume are somehow kept constant, the change in entropy from some beginning state to some final state is directly proportional to the change in potential energy.

It is then natural to ask, for a given temperature, which distribution of particles has the highest entropy of all distributions, since this would place an upper bound on the useful energy that could be extracted from a system. It is possible to show that for an ideal gas, the answer is the Maxwell-Boltzmann distribution [*Lifshitz et al.*, 1994, pp. 118-119]:

$$f(p_x, p_y, p_z) = N \left(\frac{1}{2\pi m k_B T} \right)^{3/2} \exp \left(-\frac{p_x^2 + p_y^2 + p_z^2}{2m k_B T} \right) \quad (2.36)$$

The assumptions used to derive this are modest. First, there must be some mechanism for random momentum or energy exchange between particles. In a dense gas, collisions accomplish this. In a plasma, Coulomb interactions, collisions, wave phenomena, or other processes may do the same. The second assumption is that particles are

indistinguishable. That is, if we have three particles with positions $(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ and momenta $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$, the state of affairs in hand would be indistinguishable from a set of particles with positions $(\mathbf{r}_3, \mathbf{r}_1, \mathbf{r}_2)$ and momenta $(\mathbf{p}_3, \mathbf{p}_1, \mathbf{p}_2)$. That is, there is no preferred ordering for a set of N particles. Under these assumptions, the most overwhelmingly likely state for a very large number of N ideal gas particles to be in is described by a Maxwellian velocity distribution. In some sense, the entropy is conceptually similar to a likelihood function, that is, it is a measure of the relative likelihood of a given distribution function.

The entropy per unit volume can be computed from the distribution function [[Huang, 2008](#), p. 73]:

$$S(\mathbf{r}) = -k_B \int f(\mathbf{p}, \mathbf{r}) \log f(\mathbf{p}, \mathbf{r}) d\mathbf{p} \quad (2.37)$$

The entropy of the Maxwell-Boltzmann distribution, or the equilibrium entropy, is then:

$$S_{\max}(\mathbf{r}) = -k_B N \left[\log \left(\frac{N}{(2\pi m k_B T)^{3/2}} \right) - \frac{3}{2} \right] \quad (2.38)$$

This equation can be used to estimate an upper bound on the maximum energy that is available to drive an instability but says nothing about the manner in which it might happen. We now investigate some specific cases of instabilities in plasmas.

2.3.2 Electrostatic beam instability

The electrostatic two-beam instability is one of the simplest examples of a plasma instability. The instability arises when there are two counterstreaming beams of particles, each with opposing velocities but sharing the same physical space. The reduced 1D distribution function is:

$$f(v) = N\delta(v - v_0) + N\delta(v + v_0)$$

Substituting into the electrostatic dispersion relation Equation 2.26, we have:

$$1 - \frac{\omega_p^2}{(v_0 k - \omega)^2} - \frac{\omega_p^2}{(v_0 k + \omega)^2} = 0 \quad (2.39)$$

Solving this for ω shows there is exponential growth whenever $k < \omega_p 2^{1/2} / v_0$. This result can easily be extended to a system of multiple beams:

$$f(v) = \sum_i N_i \delta(v - v_i)$$

we have:

$$1 - \frac{q^2}{\epsilon_0 m} \sum_i \frac{N_i}{(v_i k - \omega)^2} = 0 \quad (2.40)$$

This yields a large polynomial with many terms. While the roots are difficult to predict directly from the coefficients, this system is also unstable for some range of spatial frequencies k .

This instability, which exists for nearly any system of beams, has enormous practical implications for numerical simulations. We discuss this problem and its consequences for particle-in-cell (PIC) simulations in Section 6.2.3.

2.3.3 Whistler mode cyclotron instability

One commonly-encountered instability in the magnetosphere is the electron cyclotron instability. This instability leads to the growth of whistler-mode waves propagating in the magnetosphere, sometimes by many tens of decibels as a wave makes one traverse from the surface of the Earth to the opposite hemisphere along a magnetic field line.

The basic linear growth rate is given by Equation 2.32. The most important parameter from the point of view of growth is the anisotropy given by Equation 2.34. Anisotropy simply means that for a given velocity magnitude, the distribution function is not the same in every direction. A positive anisotropy means that there are more particles traveling perpendicular to the magnetic field than parallel to it. Sufficiently large positive anisotropies can lead to wave growth.

When wave amplitudes get large enough that the distribution function itself is significantly perturbed from equilibrium, predictions from linear theory can fail. In extreme cases, the assumptions used to derive the linear growth rate (i.e., smoothness and long-time behavior) break down, so it is no longer even possible to use linear theory to describe the behavior of the perturbed distribution. In the nonlinear regime, a variety of interesting effects can occur, one of the most important of which is saturation. The cyclotron instability cannot continue forever. While Equation 2.32 predicts the growth or damping rate at low amplitudes, the process must eventually saturate because the particle anisotropy is destroyed by the wave itself. Essentially, the free energy in the non-equilibrium particle distribution is given up to the wave, and when that energy is exhausted or no longer available to create an instability, the process stops.

2.3.4 Inhomogeneity-driven cyclotron instability

When spatial inhomogeneity is present, the phenomena encountered can be even more interesting. The inhomogeneity of the Earth's background magnetic field, coupled with nonlinearities induced by large-amplitude waves, is thought to be responsible for much of the characteristic behavior of chorus.

One well-developed theory of cyclotron instabilities in an inhomogeneous magnetized plasma is due to *Omura et al.* [2008], based on previous work by *Omura et al.* [1991]; *Bell* [1984, 1986]; *Nunn* [1974]. The basic theory is concerned with particle *trapping*. Sufficiently large wave fields can act as potential wells, effectively trapping particles in lockstep with the wave, thereby significantly altering the trajectories of particles near the resonant velocity:

$$v_{\text{res}} = \frac{\omega + \omega_c}{k_{\parallel}}$$

Note the sign convention: ω_c is negative for electrons, so this means that for k_{\parallel} positive, the resonant velocity is negative. In other words, the resonant electrons and the wave are traveling in opposite directions. The contribution of *Omura et al.* [1991] was to add first-order corrections to the basic resonance condition in order to write

an explicit expression for the trajectory of a particle in phase space, where the phase space angle ζ is the angle of a single particle's perpendicular velocity vector with respect to the wave magnetic field:

$$\frac{\partial^2 \zeta}{\partial t^2} = \omega_t^2 (\sin \zeta + S) \quad (2.41)$$

The trapping frequency ω_t is given by:

$$\omega_t = \sqrt{kv_{\perp} \frac{|q|B_w}{m}} \quad (2.42)$$

The term S incorporates first-order corrections to account for the spatially-varying background magnetic field and time-varying wave frequency which we write in non-relativistic form:

$$S = -\frac{1}{\omega_t^2 \delta^2} \left[\left(1 - \frac{v_{\text{res}}}{v_g}\right)^2 \frac{\partial \omega}{\partial t} + \left(\frac{kv_{\perp}^2}{2\omega_c} + \left(1 + \frac{\delta^2}{2}\right) v_{\text{res}}\right) \frac{\partial \omega_c}{\partial h} \right] \quad (2.43)$$

The term ω is the wave frequency and v_g is the wave group velocity. The dimensionless term δ is defined as $\delta^2 = 1 - \omega^2/(c^2 k^2)$. The term h is the distance along a single magnetic field line.

This equation is recognizable as the forced pendulum equation. The frequency of oscillation is roughly equal to ω_t , and the term S represents some external force acting to push the pendulum away from equilibrium. Effectively, the term S makes the pendulum oscillation lopsided, shifting its equilibrium position away from $\zeta = \pi$ and also modifying the motion away from sinusoidal.

What trapping means is more obvious when we plot the trajectories of individual particles in this system. Figure 2.4 shows some of the trajectories followed by a particle for different values of S in (ζ, v_{\parallel}) space. The term v_{\parallel} is directly related to the angular position via $\partial \zeta / \partial t = k(v_{\parallel} - v_{\text{res}})$.

The red lines indicate trapped particle trajectories. In the pendulum analogy, a trapped particle would correspond to a pendulum swinging from side to side within some small angle near its equilibrium position at $\zeta = \pi$. Blue lines denote untrapped

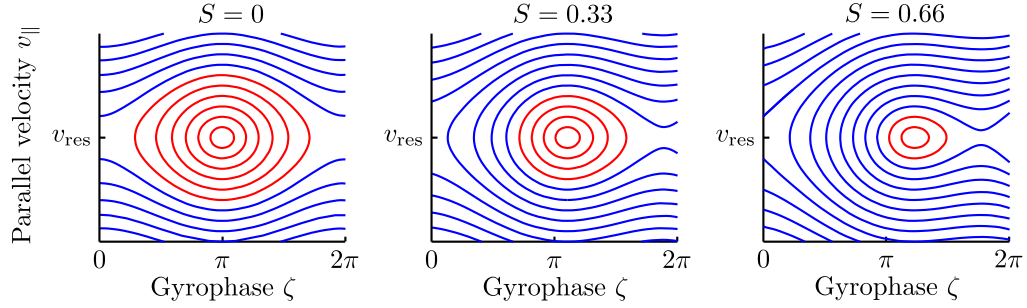


Figure 2.4: Trapped and untrapped particle trajectories in phase space. The horizontal axis is the gyrophase or angle with respect to the wave magnetic field. The vertical axis is the parallel velocity of the particle. Red indicates trapped particle trajectories while blue indicates untrapped.

particle trajectories. In the pendulum analogy, untrapped particles would correspond to the case of a pendulum pushed so hard that it rotated freely about its axis instead of swinging from side to side. The term S introduces an asymmetry in this swinging, changing both the shape and the location of the trap in phase space.

Once the particle trap is created by the wave fields, particles can neither enter nor leave the trap until the wave dissipates or the trap topology changes or is destroyed. The trap itself, to first order, is a persistent structure that can remain even if it moves into a region where the untrapped particle density is different than the density inside the trap. When this happens, the situation becomes much more complicated, because the difference in density constitutes a resonant current. If the distribution of particles inside the trap are known, then in principle, the induced resonant current can be computed by integrating over velocity space using Equation 2.8.

The asymmetry induced by S changes both the nature and the magnitude of this resonant current, so that maximum wave growth is attained when S is negative. Since S depends on the time rate of change of the wave frequency $\partial\omega/\partial t$, this implies that there will be some special class of seed waves, driven by noise, with a specific time-frequency slope for which nonlinear growth is maximized. These waves have either rising or falling frequencies depending on the sign and phase of this resonant current.

The most significant limitation of this theory is that it says nothing about the ultimate energy source that produces and amplifies these seed waves. In fact, none of these processes can happen unless the initial distribution function is sufficiently

anisotropic to drive wave growth to sufficiently high levels to initiate particle trapping. *Omura et al.* [2008] suggest that chorus is ultimately driven by noise, which is amplified by the unstable distribution function and then selectively amplified by the aforementioned process. This theory, however, makes great reliance on quasi-monochromatic behavior of the wave, that is, it is not good enough that only one trap is formed, but by making assumptions of smoothness and quasi-monochromatic waves, the theory implicitly assumes the seed waves must also be smooth, slowly-varying, and sufficiently coherent over many periods in order to produce persistent current structures that amplify the waves. We discuss more along these lines in Chapter 6, suggesting some simple mechanisms to produce coherent seed waves.

Chapter 3

The DG method

The discontinuous Galerkin (DG) method is a hybrid method for solving hyperbolic conservation-law partial differential equations. Between elements, the method resembles the finite volume method. Local to each element, the method is a *spectral method*. Spectral methods differ greatly in their details and the range of problems they are well-suited to solving. All share one common feature, that is, they all seek to use a finite-dimensional *basis* to approximate an infinite-dimensional solution space. Here, for simplicity we only consider weighted linear combinations of some predefined finite basis $\{\phi_1, \phi_2, \dots, \phi_N\} \in \Phi$, where the unknown solution $u(x)$ is approximated as:

$$u(x) \simeq \sum_j^N u_j \phi_j(x)$$

The goal is to find the weights u_j such that the approximate solution nearly matches the actual solution $u(x)$. Precisely how this is developed into a scheme for solving a specific PDE is what defines a particular spectral method.

Spectral methods are quite old, dating back to as early as Fourier's work on the heat equation. Fourier recognized that if the unknown solution to the heat equation were approximated using an orthogonal set of sinusoids, the continuous spatial variable could be integrated out, leaving behind a set of coefficients which could be solved for directly. The DG method is one of the more recent developments in solution methods of this kind. We apply it to our class of problems specifically because of its

flexibility (both geometrically and in terms of its accuracy) and because of its ease of parallelization on modern computers. We begin with a brief introduction to the method, followed by application to a plasma, paying particular attention to issues of stiffness that commonly arise in wave propagation problems at VLF in the near-Earth space environment.

Portions of this chapter have been published in the Journal of Geophysical Research as *Foust et al.* [2011b].

3.1 Description and derivation

The nodal DG formulation described in this section is based primarily on the work of *Hesthaven and Warburton* [2002] and described in the book by *Hesthaven and Warburton* [2007], and thus our notation and terminology follow theirs closely, elaborating where necessary for clarity. We first describe the DG method and then describe the minor changes necessary to make the scheme nodal.

The DG scheme begins with a conservation law:

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla \cdot \mathbf{f}(\mathbf{u}) = \mathbf{s}$$

The term \mathbf{s} is source term which may or may not depend on the unknown \mathbf{u} . Both \mathbf{u} and \mathbf{f} can be tensor fields, so that:

$$\nabla \cdot \mathbf{f} = \frac{\partial}{\partial x} f_x + \frac{\partial}{\partial y} f_y + \frac{\partial}{\partial z} f_z$$

In this notation, it is implied that if \mathbf{f} is a rank n tensor field, then the divergence $\nabla \cdot \mathbf{f}$ is of rank $n - 1$. We primarily consider problems of the form:

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla \cdot \mathbf{f}(\mathbf{u}) = \mathbf{s} \tag{3.1}$$

Here \mathbf{f} is a rank-two tensor-valued function of the vector \mathbf{u} , that is:

$$\frac{\partial u_j}{\partial t} + \frac{\partial}{\partial x_i} f_{ij} = s_j$$

We begin by dividing the continuous space into a set of discrete elements, as shown in Figure 3.1, where the mesh is chosen to conform to boundaries between any discrete objects in the domain. The element size is typically chosen such that the wavelengths of interest are adequately sampled by the grid. The choice of grid size and its effect on the convergence of the scheme are discussed more quantitatively in Section 3.5. When written in conservative form, the flux terms in Maxwell’s equations are tensor

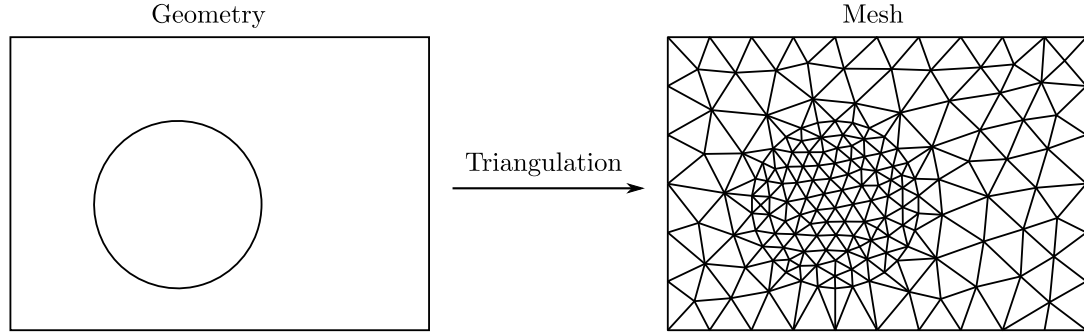


Figure 3.1: Sample illustration showing how a geometry can be discretized into a mesh of discrete triangles.

quantities. As such, we now discuss the basic method with special attention paid to the complicated tensor terms that arise. We rely heavily on overloaded notation, meaning that we often write an equation as if it were in terms of simple matrix-vector products, understanding that in fact each entry of the matrix may itself represent another matrix. We find that overloading in this manner is a useful conceptual tool when generalizing the DG method to tensor fluxes. Nevertheless, we elaborate these equations where required for clarity.

3.1.1 Derivation

We begin by approximating the unknown solution $\mathbf{u}(\mathbf{r})$ within each element as a weighted sum of basis functions:

$$\mathbf{u}^e(\mathbf{r}) \simeq \sum_{j=1}^N \mathbf{u}_j^e \phi_j^e(\mathbf{r}) \quad (3.2)$$

The functions $\phi_j(\mathbf{r})$ are the members of the basis. These could be any linearly independent functions spanning a space, e.g., polynomials or sinusoids. The weights \mathbf{u}_j

are what we seek to find. Given these weights, the approximate solution can be found by a weighted sum over the basis. The superscript e denotes that what we have is strictly an element-local expansion, that is, the basis and weights only apply within a single triangle in the mesh in Figure 3.1. Similarly, we approximate the unknown flux and source terms in the same manner:

$$\mathbf{f}^e(\mathbf{r}) \simeq \sum_{j=1}^N \mathbf{f}_j^e \phi_j^e(\mathbf{r}) \quad (3.3)$$

$$\mathbf{s}^e(\mathbf{r}) \simeq \sum_{j=1}^N \mathbf{s}_j^e \phi_j^e(\mathbf{r}) \quad (3.4)$$

Substituting the approximations into Equation 3.1, we have:

$$\frac{\partial \mathbf{u}^e(\mathbf{r})}{\partial t} + \nabla \cdot \mathbf{f}^e(\mathbf{r}) - \mathbf{s}^e(\mathbf{r}) = \mathbf{e}^e \quad (3.5)$$

The term on the right is the residual error \mathbf{e} . To convert Equation 3.5 into a scheme, we multiply the residual by each function in the basis again and set each integrated residual to zero:

$$\int_V \left(\frac{\partial \mathbf{u}^e(\mathbf{r})}{\partial t} + \nabla \cdot \mathbf{f}^e(\mathbf{r}) - \mathbf{s}^e(\mathbf{r}) \right) \phi_i^e(\mathbf{r}) d\mathbf{r} = 0, \quad i = \{1, 2, \dots, N\} \quad (3.6)$$

This method of forcing the integrated residual to zero is termed *Galerkin projection*, and ensures that the residual is orthogonal to the space spanned by the basis. For linear problems, this means that the approximate solution expanded in the given basis is as close as possible to the true solution in a least-squares sense [Roman, 2005, pp. 315-316]. This concept is illustrated in Figure 3.2, where some basis in \mathbb{R}^1 spans a one-dimensional subset of the full two-dimensional space of residuals in \mathbb{R}^2 . The true solution is where the residual error goes to zero, that is, at the origin. The approximate solution found by Galerkin projection is the point where the residual error is orthogonal to the space spanned by the basis. This simple 2D figure only illustrates the basic concept of Galerkin projection. In reality, our true solution exists in \mathbb{R}^∞ and we seek to approximate it in some smaller space \mathbb{R}^N using

an N -dimensional basis.

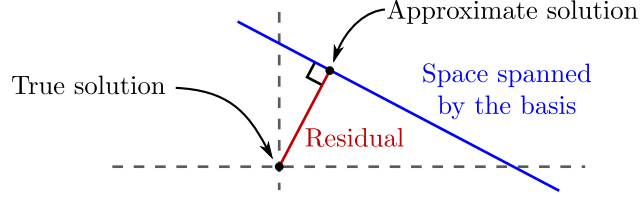


Figure 3.2: Conceptual illustration of Galerkin projection. The 2D space is the space of all possible residuals. The true solution is found where the residual goes to zero (the origin). The approximate solution expanded in some 1D basis spans a 1D subset of this space, indicated in blue. Galerkin projection picks the approximate solution such that the error residual (indicated in red) is orthogonal to the space spanned by the basis.

At this point, the boundary conditions on each element are still unspecified, so the solution cannot possibly be unique. In the DG method, this problem is solved by borrowing an idea from the finite volume technique. For conservation laws, the volume integral of the divergence of the flux over one element is related to the surface integral of the flux on the boundary of that element, a relation known as Gauss's theorem:

$$\int_V \nabla \cdot \mathbf{f} \, d\mathbf{r} = \oint_S \hat{\mathbf{n}} \cdot \mathbf{f} \, d\mathbf{r} \quad (3.7)$$

Integrating Equation 3.6 by parts once (applying Gauss's theorem), we have the following, where we have dropped the superscript e , understanding that this is still an element-local statement:

$$\int_V \left(\frac{\partial \mathbf{u}(\mathbf{r})}{\partial t} - \mathbf{f}(\mathbf{r}) \cdot \nabla - \mathbf{s}(\mathbf{r}) \right) \phi_i(\mathbf{r}) \, d\mathbf{r} = - \oint_S (\hat{\mathbf{n}} \cdot \mathbf{f}^*(\mathbf{r})) \phi_i(\mathbf{r}) \, d\mathbf{r} \quad (3.8)$$

On the right-hand side, we have replaced the flux \mathbf{f} with an unknown quantity \mathbf{f}^* called the *numerical flux*, a concept borrowed from the finite volume technique. The numerical flux is essentially a well-educated estimate of the unknown flux at the interface between two elements, chosen such that it combines information interior to the element and exterior to the element in a physically consistent manner.

Equation 3.8 is the DG scheme in so-called *weak form*. The term weak here is used loosely and refers to the fact that the boundary conditions depend only on the

numerical flux. Integrating Equation 3.8 by parts again, we have the DG scheme in its *strong form*, where the right-hand side includes both a local flux term \mathbf{f} and the unknown numerical flux \mathbf{f}^* :

$$\int_V \left(\frac{\partial \mathbf{u}(\mathbf{r})}{\partial t} + \nabla \cdot \mathbf{f}(\mathbf{r}) - \mathbf{s}(\mathbf{r}) \right) \phi_i(\mathbf{r}) d\mathbf{r} = \oint_S \hat{\mathbf{n}} \cdot (\mathbf{f}(\mathbf{r}) - \mathbf{f}^*(\mathbf{r})) \phi_i(\mathbf{r}) d\mathbf{r} \quad (3.9)$$

Recalling that we have approximated the unknown quantities in a basis expansion using Equations 3.2, 3.3, and 3.4, we see that the only terms that depend on \mathbf{r} are the members of the basis. Thus we can evaluate the integrals for each pair (ϕ_i, ϕ_j) in the basis, yielding the matrix version of the weak form:

$$\mathbf{M} \frac{d\mathbf{u}}{dt} - \mathbf{K}^T \cdot \mathbf{f} - \mathbf{M}\mathbf{s} = -\mathbf{L} (\hat{\mathbf{n}} \cdot \mathbf{f}^*) \quad (3.10)$$

and the strong form:

$$\mathbf{M} \frac{d\mathbf{u}}{dt} + \mathbf{K} \cdot \mathbf{f} - \mathbf{M}\mathbf{s} = \mathbf{L} (\hat{\mathbf{n}} \cdot \mathbf{f} - \hat{\mathbf{n}} \cdot \mathbf{f}^*) \quad (3.11)$$

The mass, stiffness, and surface mass matrices, (\mathbf{M} , \mathbf{K} , and \mathbf{L} , respectively), are defined as follows, where the surface mass matrix \mathbf{L} is evaluated only along the faces of the element (denoted as a subscript F in the integral):

$$M_{ij} = \int_V \phi_i(\mathbf{r}) \phi_j(\mathbf{r}) d\mathbf{r} \quad (3.12a)$$

$$K_{ij} = \int_V \phi_i(\mathbf{r}) \nabla \phi_j(\mathbf{r}) d\mathbf{r} \quad (3.12b)$$

$$L_{ij} = \int_F \phi_i(\mathbf{r}) \phi_j(\mathbf{r}) d\mathbf{r} \quad (3.12c)$$

We are now heavily overloading notation, so some clarification is in order. The original unknown $\mathbf{u}(\mathbf{r})$ is a continuous vector-valued function of spatial coordinates. For the sake of illustration, we will assume it is a directed vector \mathbb{R}^3 , i.e., $\mathbf{r} = (x, y, z)$ and $\mathbf{u}(\mathbf{r}) = [u_x(\mathbf{r}), u_y(\mathbf{r}), u_z(\mathbf{r})]$. After discretization, each continuous component of this vector is converted into a vector of unknown weights of finite length N , where N is the number of functions in the basis. Therefore, we now have the vector of

vectors $(\mathbf{u}_x, \mathbf{u}_y, \mathbf{u}_z)$. The term $\mathbf{M}\mathbf{u}$, then, should be understood to mean the block-componentwise multiplication of the $N \times N$ matrix \mathbf{M} by each of these N -length vectors, that is:

$$\mathbf{M}\mathbf{u} = \begin{bmatrix} \mathbf{M}\mathbf{u}_x \\ \mathbf{M}\mathbf{u}_y \\ \mathbf{M}\mathbf{u}_z \end{bmatrix}$$

The stiffness tensor \mathbf{K} is most easily thought of as three $N \times N$ matrices, one for each component of the gradient, that is, $\mathbf{K} = (\mathbf{K}_x, \mathbf{K}_y, \mathbf{K}_z)$. Thus, the dot product term in Equation 3.11 should be understood as:

$$\mathbf{K} \cdot \mathbf{f} = \mathbf{K}_x \mathbf{f}_x + \mathbf{K}_y \mathbf{f}_y + \mathbf{K}_z \mathbf{f}_z = \begin{bmatrix} \mathbf{K}_x \mathbf{f}_{xx} + \mathbf{K}_y \mathbf{f}_{yx} + \mathbf{K}_z \mathbf{f}_{zx} \\ \mathbf{K}_x \mathbf{f}_{xy} + \mathbf{K}_y \mathbf{f}_{yy} + \mathbf{K}_z \mathbf{f}_{zy} \\ \mathbf{K}_x \mathbf{f}_{xz} + \mathbf{K}_y \mathbf{f}_{yz} + \mathbf{K}_z \mathbf{f}_{zz} \end{bmatrix}$$

Similarly:

$$\hat{\mathbf{n}} \cdot \mathbf{f} = n_x \mathbf{f}_x + n_y \mathbf{f}_y + n_z \mathbf{f}_z = \begin{bmatrix} n_x \mathbf{f}_{xx} + n_y \mathbf{f}_{yx} + n_z \mathbf{f}_{zx} \\ n_x \mathbf{f}_{xy} + n_y \mathbf{f}_{yy} + n_z \mathbf{f}_{zy} \\ n_x \mathbf{f}_{xz} + n_y \mathbf{f}_{yz} + n_z \mathbf{f}_{zz} \end{bmatrix}$$

The surface normal $\hat{\mathbf{n}}$ is the outward normal on each face of a triangle (in 2D) or tetrahedron (in 3D).

The numerical flux term \mathbf{f}^* provides a mechanism to exchange some conserved quantity (an integrated flux) between elements, thus ensuring a globally valid solution. However, in contrast to the finite element method, we do not explicitly require continuity of the fields at the boundaries between elements. In other words, the solution is discontinuous. This property may seem, at first glance, to be a disadvantage of the method, but it in fact has some major advantages. First, the scheme is highly parallelizable. Unlike in the finite element method, the mass and stiffness matrices are local and small, so no large matrix inversion is required. Second, the scheme has extremely robust handling of material discontinuities, since each element-local

expansion is completely independent of its neighbors.

3.1.2 Nodal DG formulation

The nodal DG method described here is due to *Hesthaven and Warburton* [2002]. The basic idea is to replace the basis described in the previous section with an interpolating basis. In other words, we define a set of interpolation points on each triangle and use the resulting interpolating polynomials to approximate the solution. The overall picture is illustrated in Figure 3.3.

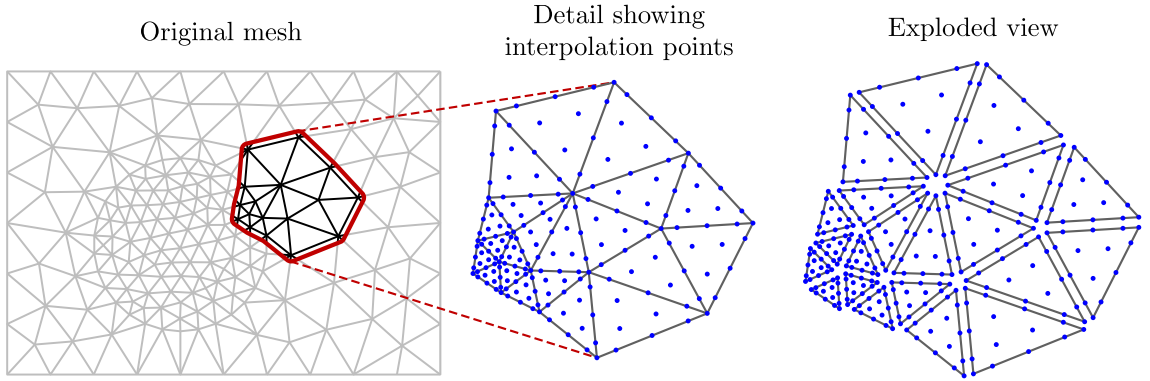


Figure 3.3: Exploded view demonstrating the nodal DG formulation. A set of interpolation control points are defined on each triangle, as shown in the middle panel. The interpolation is independent on each triangle, as shown in the rightmost view.

Constructing the basis in this way leads to a very intuitive representation of the approximate solution. The values at the control points are the solution at those points. The values anywhere else in the element can be found by simply interpolating using the element's basis and control points. In addition, the scheme simplifies the construction of the surface integral terms. Since only points defined on the faces of each element contribute to the surface integral there, we can rewrite the surface integral matrix \mathbf{L} as a reduced sum over each face. We can then rewrite the surface integral term as a sum over the contributions from each face:

$$\mathbf{L}(\hat{\mathbf{n}} \cdot \mathbf{f}) = \sum_{k=1}^{N_{\text{faces}}} \mathbf{L}^k(\hat{\mathbf{n}}^k \cdot \mathbf{f})$$

The superscript k denotes a face of the element, which would be each end of a line

segment in 1D, each edge of a triangle in 2D, and each face of a tetrahedron in 3D. Each face matrix \mathbf{L}^k can be written as a non-square matrix, since only the points on each face contribute to the surface integral there. This rewriting is accomplished by changing the second index j in Equation 3.12c to include only points on that face.

The order N_p refers to the order of the interpolating polynomials, which in turn determines the number of points defined within each triangle, as illustrated in Figure 3.4 for polynomial orders 1, 3, and 5. The nodal sets are chosen to optimize the quality of the interpolation over a given triangle, and thus are more clustered near the corners of the elements.

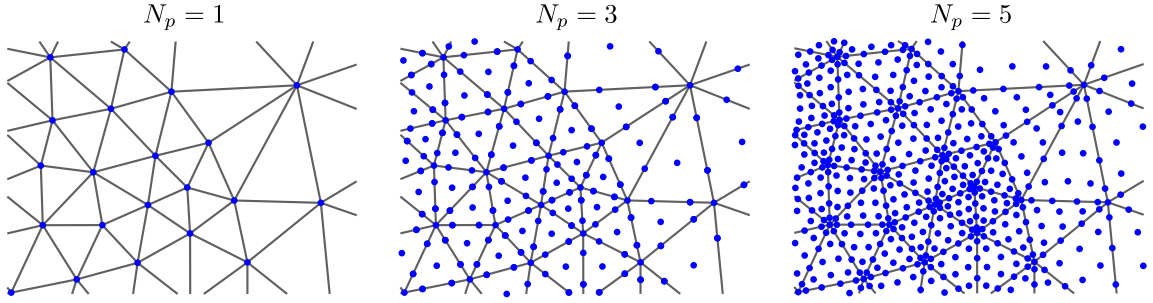


Figure 3.4: Diagram showing the node locations for a sample mesh for polynomial orders N_p equal to 1, 3, and 5. First order is linear interpolation, requiring only three points to specify a plane. Orders 3 and 5 incorporate more points, including points on the interior of the element.

3.1.3 Application to Maxwell's equations

The nodal DG method described in the previous sections can be applied easily to Maxwell's equations. We only need to write Maxwell's equations in flux conservation form and then identify a suitable numerical flux. Maxwell's equations are:

$$\frac{\partial \mathbf{D}}{\partial t} - \nabla \times \mathbf{H} + \mathbf{J} = 0 \quad (3.13a)$$

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} + \mathbf{J}_m = 0 \quad (3.13b)$$

Coupled with the constitutive relations:

$$\mathbf{D} = \vec{\epsilon} \mathbf{E} \quad (3.14a)$$

$$\mathbf{B} = \vec{\mu} \mathbf{H} \quad (3.14b)$$

Note that we include a magnetic current \mathbf{J}_m . Although magnetic currents do not exist in nature, it is often more convenient to treat a complicated material as if it had an equivalent magnetic current instead of modeling the material via the constitutive relation. Rewriting in conservative form, we have:

$$\frac{\partial \mathbf{D}}{\partial t} + \nabla \cdot \mathbf{f}_E + \mathbf{J} = 0 \quad (3.15a)$$

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \cdot \mathbf{f}_H + \mathbf{J}_m = 0, \quad (3.15b)$$

where the tensor fluxes \mathbf{f}_E and \mathbf{f}_H are defined as:

$$\mathbf{f}_E = \begin{bmatrix} 0 & H_z & -H_y \\ -H_z & 0 & H_x \\ H_y & -H_x & 0 \end{bmatrix} \quad \mathbf{f}_H = \begin{bmatrix} 0 & -E_z & E_y \\ E_z & 0 & -E_x \\ -E_y & E_x & 0 \end{bmatrix} \quad (3.16)$$

so that:

$$\nabla \cdot \mathbf{f}_E = -\nabla \times \mathbf{H} \quad \nabla \cdot \mathbf{f}_H = \nabla \times \mathbf{E}$$

The DG scheme follows directly from the results in the previous section. In strong form:

$$\mathbf{M} \frac{d(\epsilon \mathbf{E})}{dt} + \mathbf{K} \cdot \mathbf{f}_E + \mathbf{M} \mathbf{J} = \mathbf{L} (\hat{\mathbf{n}} \cdot \mathbf{f}_E - \hat{\mathbf{n}} \cdot \mathbf{f}_E^*) \quad (3.17a)$$

$$\mathbf{M} \frac{d(\mu \mathbf{H})}{dt} + \mathbf{K} \cdot \mathbf{f}_H + \mathbf{M} \mathbf{J}_m = \mathbf{L} (\hat{\mathbf{n}} \cdot \mathbf{f}_H - \hat{\mathbf{n}} \cdot \mathbf{f}_H^*) \quad (3.17b)$$

We use a standard numerical flux [[Hesthaven and Warburton, 2002](#)]:

$$\hat{\mathbf{n}} \cdot (\mathbf{f}_E - \mathbf{f}_E^*) = \frac{1}{(Z^+ + Z^-)} (\alpha \hat{\mathbf{n}} (\hat{\mathbf{n}} \cdot [\mathbf{E}]) - \alpha [\mathbf{E}] - Z^+ \hat{\mathbf{n}} \times [\mathbf{H}]) \quad (3.18a)$$

$$\hat{\mathbf{n}} \cdot (\mathbf{f}_H - \mathbf{f}_H^*) = \frac{1}{(Y^+ + Y^-)} (\alpha \hat{\mathbf{n}} (\hat{\mathbf{n}} \cdot [\mathbf{H}]) - \alpha [\mathbf{H}] + Y^+ \hat{\mathbf{n}} \times [\mathbf{E}]) \quad (3.18b)$$

The bracketed notation $[\mathbf{u}]$ means the field difference at the face of a given element, $[\mathbf{u}] = \mathbf{u}^- - \mathbf{u}^+$, where \mathbf{u}^- denotes the field values at a face interior to a given element,

and \mathbf{u}^+ denotes the field values at the face of the element's immediate neighbor. Z is the impedance, $(\mu/\epsilon)^{1/2}$. Y is the admittance, $(\epsilon/\mu)^{1/2}$. The term α is variable. If $\alpha=0$, the flux is a so-called *central flux*, essentially an average of the field values weighted by the harmonic average of the characteristic impedance or admittance. If $\alpha=1$, the flux is an *upwinding flux*, which attempts to take into account the direction of propagation of the various characteristic modes of the system. The derivation of the upwind flux is given in Appendix B.

Whether upwind fluxes or central fluxes should be used depends greatly on the problem in hand. Upwind fluxes tend to be mildly dissipative, but this dissipation can be an advantage in suppressing high-frequency noise. Central fluxes never exhibit dissipation, but instead unresolved modes can have negative group velocities. Resolution problems, then, can be more difficult to spot when using central fluxes, since energy is perfectly preserved even when the solution is very inaccurate. However, central fluxes can better tolerate a complete loss of hyperbolicity. We discuss this problem further in Section 3.3, when we describe the spectral properties of our cold plasma DG scheme. In practice, either upwind fluxes or central fluxes can be used for many problems. Either choice yields nearly identical answers except in pathological cases.

3.2 Linear dispersive media

We now describe our approach to incorporate arbitrary linear media in the DG framework. We begin by rewriting Maxwell's equations to collect all frequency dependence and nonlinearities in the constitutive relations as effective electric and magnetic currents \mathbf{J} and \mathbf{J}_m :

$$\epsilon_\infty \frac{\partial \mathbf{E}}{\partial t} - \nabla \times \mathbf{H} + \mathbf{J} \left(\frac{\partial \mathbf{E}}{\partial t}, \mathbf{E}, t \right) = 0 \quad (3.19a)$$

$$\mu_\infty \frac{\partial \mathbf{H}}{\partial t} + \nabla \times \mathbf{E} + \mathbf{J}_m \left(\frac{\partial \mathbf{H}}{\partial t}, \mathbf{H}, t \right) = 0 \quad (3.19b)$$

Here we have introduced the simple scalar terms ϵ_∞ and μ_∞ , which represent the (assumed isotropic) scalar permittivity and permeability at infinite frequency. On

physical grounds, we know that as the frequency tends to infinity, the terms ϵ and μ approach their free space values ϵ_0 and μ_0 , which are simple non-frequency dependent scalars. We can therefore simply subtract off the term at infinity, which leaves us with the non-frequency dependent terms ϵ_∞ and μ_∞ in front of the time derivatives. In other words, we write the frequency-dependent permittivity and permeability as a sum of a constant term (ϵ_∞ and μ_∞) and a frequency-dependent equivalent conductivity (see Equation 3.21). Collecting all of the remaining terms yields the expressions for the equivalent currents \mathbf{J} and \mathbf{J}_m . After DG discretization, we are left with a differential algebraic equation (DAE) of the form:

$$f\left(\frac{d\mathbf{u}}{dt}, \mathbf{u}, t\right) = 0$$

Equations like this are usually solved iteratively because of the implicit dependence on the time derivatives of the field. Solution methods for DAEs are typically much slower than those for simple explicit ODEs but DAEs can, in principle, be solved. However, we do not investigate the solution of DAEs since cold plasmas and other common types of materials do not have this kind of implicit dependence. We therefore only consider forms such as:

$$\epsilon_\infty \frac{\partial \mathbf{E}}{\partial t} - \nabla \times \mathbf{H} + \mathbf{J}(\mathbf{E}, t) = 0 \quad (3.20a)$$

$$\mu_\infty \frac{\partial \mathbf{H}}{\partial t} + \nabla \times \mathbf{E} + \mathbf{J}_m(\mathbf{H}, t) = 0, \quad (3.20b)$$

which lead to explicit ODEs of the form:

$$\frac{d\mathbf{u}}{dt} = f(\mathbf{u}, t)$$

Explicit ODEs of this form can be solved easily using any of a number of well-developed techniques. All of the materials we consider in this dissertation can be described this way, but we emphasize that we need not require linearity of the equivalent currents. In this section, we consider only linear dispersive media, but the idea can be easily applied to nonlinear media with the appropriate expression for the

current. In Chapter 4, for example, we discuss our incorporation of nonlinear plasma behavior using the particle-in-cell (PIC) method.

Based on the discussion above, we can rewrite any linear permittivity as an equivalent conductivity using the following relation:

$$\vec{\epsilon}(\omega) = \epsilon_\infty \mathbf{I} - \frac{\vec{\sigma}(\omega)}{j\omega}, \quad (3.21)$$

where \mathbf{I} is the unit dyad (the identity matrix). We explicitly denote ϵ_∞ as the permittivity at infinite frequency, although for any physical system, $\epsilon_\infty = \epsilon_0$ as $\omega \rightarrow \infty$. However, it is sometimes useful for computational reasons to be able to choose different values.

By incorporating all of the frequency-dependent behavior in the equivalent current, we need not change the basic DG formulation in order to handle different types of materials. Further, the approach is natural. The nodal DG approach solves for the field values at interpolation points, so the relationship between the current and the causative electric field at those points is direct. The electric field determines the current through some pointwise set of ordinary differential equations, described in the frequency domain as:

$$\mathbf{J}(s) = \vec{\sigma}(s) \mathbf{E}(s),$$

where we are now using s to denote the complex frequency instead of $-j\omega$. This formulation is particularly useful for a plasma since the cold plasma conductivity tensors sum for each component species:

$$\mathbf{J}(s) = (\vec{\sigma}_1(s) + \vec{\sigma}_2(s) + \vec{\sigma}_3(s) + \dots) \mathbf{E}(s)$$

Typically either the permittivity tensor $\vec{\epsilon}(s)$ or the conductivity tensor $\vec{\sigma}(s)$ are specified in frequency-domain form. For a linear material, each entry in the tensor is a simple ratio of polynomials in s . Rather than developing an ad-hoc time-stepping approach specifically tailored to each type of material, it is more flexible to utilize the already very large body of highly accurate techniques to solve first-order ODEs

of the form:

$$\frac{d\mathbf{u}}{dt} = f(\mathbf{u})$$

The task of converting such a frequency-domain description into a first-order system is a well-known problem from control literature termed *system realization*. We now describe our approach and its application to two types of linear, dispersive media.

3.2.1 Realization

Our goal is to convert a frequency-domain description relating the current to the causative electric field $\mathbf{J}(s) = \vec{\sigma}(s)\mathbf{E}(s)$ to a first-order ODE, which we write as:

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{E} \quad (3.22a)$$

$$\mathbf{J} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{E} \quad (3.22b)$$

Each of these matrices has a useful interpretation. \mathbf{A} describes the dynamics of the system. The characteristic frequencies are the eigenvalues of this matrix. \mathbf{B} describes how the input \mathbf{E} maps into the current state. \mathbf{C} describes how the state \mathbf{x} maps into the output \mathbf{J} . \mathbf{D} , the pass-through matrix, dominates as the frequency approaches infinity.

A realization is correct if the output of this ODE has the same transient and time-harmonic response as the frequency-domain description for all possible inputs. Realizations, however, are not unique. In fact, there may be an infinite number of matrices $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ consistent with the given frequency-domain description, and there exist many published algorithms to compute them [*Ho and Kalman, 1966*; *De Schutter and De Moor, 1995*; *De Schutter, 2000*]. However, not all realizations are equally well-suited for numerical simulation.

A given frequency-domain description has a finite number of internal degrees of freedom. A realization is termed *minimal* if the state vector \mathbf{x} is only as large as needed to represent all the degrees of freedom. If the state vector is larger, this may mean that there are hidden states or unnecessary, redundant states. In a computational setting,

a non-minimal realization wastes both memory and computing time. In addition, for numerical stability it is essential that the matrices are all well-conditioned so that the inputs do not map unequally into the current state and the current state does not map unequally into the output. In either case, finite precision effects in an ill-conditioned system could lead to excessive numerical noise or even instability.

The technique we use is a SVD-based (singular value decomposition) balanced minimal realization technique due to [Antsaklis \[1997, p. 423\]](#). The first step is to find the minimal polynomial $m(s)$ of the system, defined as the least common denominator of every entry $\sigma_{ij}(s)$ of the conductivity tensor. The integer r is the order of this resulting polynomial. We first expand the conductivity tensor in a series:

$$\vec{\sigma}(s) = \vec{\sigma}_0 + \vec{\sigma}_1 s^{-1} + \vec{\sigma}_2 s^{-2} + \dots \quad (3.23)$$

The matrices $\vec{\sigma}_0, \vec{\sigma}_1, \vec{\sigma}_2, \dots$ are called the *Markov parameters* of the system, which are defined as:

$$\begin{aligned} \vec{\sigma}_0 &= \lim_{s \rightarrow \infty} \vec{\sigma}(s) \\ \vec{\sigma}_1 &= \lim_{s \rightarrow \infty} s(\vec{\sigma}(s) - \vec{\sigma}_0) \\ \vec{\sigma}_2 &= \lim_{s \rightarrow \infty} s^2(\vec{\sigma}(s) - \vec{\sigma}_0 - \vec{\sigma}_1 s^{-1}) \\ &\dots \end{aligned} \quad (3.24)$$

We then define the block matrices \mathbf{G} and $\tilde{\mathbf{G}}$:

$$\mathbf{G} = \begin{bmatrix} \vec{\sigma}_1 & \cdots & \vec{\sigma}_r \\ \vdots & \ddots & \vdots \\ \vec{\sigma}_r & \cdots & \vec{\sigma}_{2r-1} \end{bmatrix} \quad (3.25)$$

$$\tilde{\mathbf{G}} = \begin{bmatrix} \vec{\sigma}_2 & \cdots & \vec{\sigma}_{r+1} \\ \vdots & \ddots & \vdots \\ \vec{\sigma}_{r+1} & \cdots & \vec{\sigma}_{2r} \end{bmatrix} \quad (3.26)$$

The tensor $\vec{\sigma}$ has dimensions 3×3 , so the matrices \mathbf{G} and $\tilde{\mathbf{G}}$ have dimensions $3r \times 3r$. We then compute the SVD of the matrix \mathbf{G} , yielding:

$$\mathbf{G} = \begin{bmatrix} \mathbf{U} & \mathbf{U}_N \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V} & \mathbf{V}_N \end{bmatrix}^T \quad (3.27)$$

The matrices \mathbf{U}_N and \mathbf{V}_N are the nullspace singular vectors, which we have written explicitly as separate matrices because we will discard them. The matrices \mathbf{U} and \mathbf{V} are the left and right singular vectors, respectively, and Σ is the diagonal matrix of singular values. We can equivalently write \mathbf{G} as:

$$\mathbf{G} = \mathbf{U} \Sigma \mathbf{V}^T \quad (3.28)$$

The number of nonzero singular values are the number of degrees of freedom in the system. The realization is then computed by:

$$\mathbf{A} = \left(\Sigma^{-\frac{1}{2}} \mathbf{U}^T \right) \tilde{\mathbf{G}} \left(\mathbf{V} \Sigma^{-\frac{1}{2}} \right) \quad (3.29a)$$

$$\mathbf{B} = \left(\Sigma^{\frac{1}{2}} \mathbf{V}^T \right) \begin{bmatrix} \mathbf{I}_{3 \times 3} & \mathbf{0}_{3 \times (3r-3)} \end{bmatrix}^T \quad (3.29b)$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{I}_{3 \times 3} & \mathbf{0}_{3 \times (3r-3)} \end{bmatrix} \left(\mathbf{U} \Sigma^{\frac{1}{2}} \right) \quad (3.29c)$$

$$\mathbf{D} = \vec{\sigma}_0 \quad (3.29d)$$

The matrices \mathbf{I} and $\mathbf{0}$ are the identity matrix and all-zeros matrix, respectively. The subscripts denote their dimensions. Supposing the number of nonzero singular values is m , then the matrices are sized: $\mathbf{A} \in \mathbb{R}^{m \times m}$, $\mathbf{B} \in \mathbb{R}^{m \times 3}$, $\mathbf{C} \in \mathbb{R}^{3 \times m}$, and $\mathbf{D} \in \mathbb{R}^{3 \times 3}$.

This realization is both minimal and balanced. Minimal means that the state vector \mathbf{x} and state transition matrix \mathbf{A} are as small as possible, that is, there are only enough states as needed to capture the dynamics of the system. Balanced means that the realization is equally observable and controllable [[Gilbert, 1963](#)]. These properties lead to well-conditioned matrices with good stability and noise properties. However, we acknowledge that other realizations may have more numerically desirable properties such as more sparsity in the state transition or mixing matrices. However,

since an existing realization can be transformed into many different equivalent forms by coordinate rotations or similarity transforms, finding only one realization suffices. Other forms can be found by simple matrix operations in a post-processing step.

3.2.2 Application to a cold plasma

The cold plasma conductivity tensor in frequency domain form is given by Equation 2.19. The minimal polynomial of this system can be computed exactly:

$$m(s) = (s + \nu)(s^2 + 2s\nu + \nu^2 + \omega_c^2) = d(s) \quad (3.30)$$

The procedure follows directly, with one caveat: we have found that when forming the Hankel matrix for a typical plasma, the entries can become extremely large, leading to severe ill-conditioning. The end result is that the zero singular values are not clearly delineated from the non-zero singular values, leading to incorrect realizations with far too many or far too few states. A solution to this ill-conditioning is to pre-scale the frequency s , i.e., $\tilde{s} = s/k$. We choose k such that the roots of the characteristic polynomial are as close to one as possible. We have found that setting k equal to the largest absolute value of the roots of the minimal polynomial $m(s)$ is a good heuristic procedure. After carrying out the realization procedure, we are left with temporary scaled matrices \tilde{A} , \tilde{B} , \tilde{C} , and \tilde{D} . The true matrices can be found by scaling back as: $A = k\tilde{A}$, $B = k\tilde{B}$, $C = \tilde{C}$, and $D = \tilde{D}$.

3.2.3 Application to the PML

An absorbing or radiation boundary condition is typically required for time-domain simulations on a truncated domain. Without it, waves numerically reflect from the boundary and corrupt the solution. The PML, or perfectly matched layer, is one type of absorbing boundary. It is a theoretically perfect medium in that it admits waves with identically zero reflection and then damps them as they propagate further into the absorbing medium. The PML was originally discovered by [Bérenger \[1994\]](#) and formulated in a rather ad-hoc way using a nonphysical field splitting. It was later shown that the PML was equivalent to either a complex stretching transform on the

space [Chew and Weedon, 1994] or a material with an anisotropic tensor constitutive relation [Gedney, 1996]. The two formulations are equivalent in terms of their wave behavior but have different physical laws and numerical behavior.

The space stretching PML is equivalent to the original PML described by Bérenger [1994] and nearly all other PML formulations developed since [Roden and Gedney, 2000; Chevalier and Inan, 2004]. Formally, the method analytically continues the spatial domain, allowing the coordinates themselves to take on complex values. The tensor form of the PML redefines the constitutive relations using frequency-dependent metric transforms on the permittivity and permeability. This form includes the UPML [Gedney, 1996] and a PML for bianisotropic media [Teixeira and Chew, 1998].

The space stretching PML is popular due to its relative ease of implementation. Since only the curl operators are affected in this PML formulation, the PML is explicitly medium-independent and thus can be adapted to any other type of medium with ease. With our realization approach, however, the tensor form presents no particular difficulty and indeed has some advantages. First, we can avoid modifying either the time-stepping scheme or our DG scheme itself. Second, the PML can be treated as just another type of material, entirely characterized by some equivalent magnetic and electric conductivity and thus fundamentally no more complicated or different than any other linear, dispersive material.

Teixeira and Chew [1998] showed how the tensor form of the PML can be constructed for any type of dispersive, anisotropic material. The end form closely resembles a metric transformation of the material tensors $\vec{\epsilon}(s)$ and $\vec{\mu}(s)$:

$$\vec{\epsilon}_{\text{pml}}(s) = \frac{1}{\det \mathbf{S}} \mathbf{S} \vec{\epsilon}(s) \mathbf{S} \quad (3.31a)$$

$$\vec{\mu}_{\text{pml}}(s) = \frac{1}{\det \mathbf{S}} \mathbf{S} \vec{\mu}(s) \mathbf{S} \quad (3.31b)$$

The matrix \mathbf{S} is defined as:

$$\mathbf{S} = \begin{bmatrix} \eta_x^{-1} & 0 & 0 \\ 0 & \eta_y^{-1} & 0 \\ 0 & 0 & \eta_z^{-1} \end{bmatrix} \quad (3.32)$$

The stretching parameters η_x , η_y , and η_z are typically complex functions of the frequency s , chosen so that waves are damped in the direction normal to the PML interface. If the normal to the boundary is $\hat{\mathbf{n}} = (1, 0, 0)$, we set the PML stretching parameters to $\boldsymbol{\eta} = (\eta, 0, 0)$. In a corner region where two faces meet, we set two of the stretching parameters, e.g., $\boldsymbol{\eta} = (\eta, \eta, 0)$ for an x - y corner. Where three faces meet, $\boldsymbol{\eta} = (\eta, \eta, \eta)$. We have illustrated this in Figure 3.5.

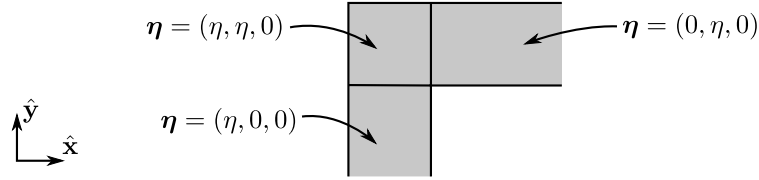


Figure 3.5: Illustration of a PML showing how the stretching parameter $\boldsymbol{\eta}$ is defined in each of the regions.

We first illustrate the construction of the tensor PML for an isotropic, non-dispersive medium. In this case, the material reduces to the UPML [Gedney, 1996]:

$$\vec{\epsilon}_{\text{pml}}(s) = \epsilon \tilde{\mathbf{S}}(s) \quad (3.33a)$$

$$\vec{\mu}_{\text{pml}}(s) = \mu \tilde{\mathbf{S}}(s) \quad (3.33b)$$

where:

$$\tilde{\mathbf{S}} = \begin{bmatrix} \eta_x^{-1} \eta_y \eta_z & 0 & 0 \\ 0 & \eta_x \eta_y^{-1} \eta_z & 0 \\ 0 & 0 & \eta_x \eta_y \eta_z^{-1} \end{bmatrix} \quad (3.34)$$

It may be tempting to attempt to generalize this to interfaces not aligned along the primary axes, e.g., we may attempt to truncate a boundary with face normal $\hat{\mathbf{n}} = (\cos 45^\circ, \sin 45^\circ, 0)$ by setting $\boldsymbol{\eta} = \eta(\cos 45^\circ, \sin 45^\circ, 0)$, but this approach is incorrect. The resulting PML is a corner PML and will not correctly truncate a face aligned at 45 degrees. The correct approach is to first construct a PML aligned along one of the primary axes and then rotate the resulting material tensor into a new reference frame, as we show below.

We use the complex frequency-shifted PML (CFS-PML) described by [Gedney and Zhao \[2010\]](#). In our notation, ξ_i is the PML loss term and α_i is a small term chosen to make the PML strictly causal. A nonzero α_i eliminates late-time reflections and other problems related to the weak causality of the original UPML formulation. The subscript i refers to one of the principal directions x , y , or z .

$$\eta_i = 1 + \frac{\xi_i}{\alpha_i + s\epsilon_0} = \frac{s + \left(\frac{\alpha_i + \xi_i}{\epsilon_0}\right)}{s + \frac{\alpha_i}{\epsilon_0}} \quad (3.35)$$

To simplify notation, we define $\beta_i = (\alpha_i + \xi_i)/\epsilon_0$ and $\gamma_i = \alpha_i/\epsilon_0$. Thus,

$$\eta_i = \frac{s + \beta_i}{s + \gamma_i} \quad (3.36)$$

The equivalent electric and magnetic conductivities are then:

$$\vec{\sigma}_{\text{pml}}(s) = \vec{\epsilon}s [\tilde{\mathbf{S}}(s) - \mathbf{I}] \quad (3.37a)$$

$$\vec{\sigma}_{m,\text{pml}}(s) = \vec{\mu}s [\tilde{\mathbf{S}}(s) - \mathbf{I}] \quad (3.37b)$$

We write the term $s(\tilde{\mathbf{S}} - \mathbf{I})$ as a new matrix $\mathbf{G}(s)$. Explicitly:

$$\mathbf{G}(s) = \begin{bmatrix} g_{xx}(s) & 0 & 0 \\ 0 & g_{yy}(s) & 0 \\ 0 & 0 & g_{zz}(s) \end{bmatrix} \quad (3.38)$$

where:

$$\begin{aligned} g_{xx}(s) &= s \left[\frac{(s + \gamma_x)(s + \beta_y)(s + \beta_z)}{(s + \beta_x)(s + \gamma_y)(s + \gamma_z)} - 1 \right] \\ g_{yy}(s) &= s \left[\frac{(s + \beta_x)(s + \gamma_y)(s + \beta_z)}{(s + \gamma_x)(s + \beta_y)(s + \gamma_z)} - 1 \right] \\ g_{zz}(s) &= s \left[\frac{(s + \beta_x)(s + \beta_y)(s + \gamma_z)}{(s + \gamma_x)(s + \gamma_y)(s + \beta_z)} - 1 \right] \end{aligned}$$

We have assumed so far that the PML is being used to absorb outgoing waves propagating in a simple medium, but the technique can be easily extended to handle anisotropic, dispersive media. Suppose we have a tensor conductivity composed of a sum of constituent conductivity tensors:

$$\vec{\sigma}(s) = \vec{\sigma}_1(s) + \vec{\sigma}_2(s) + \dots$$

From Equation 3.21, the equivalent permittivity is:

$$\vec{\epsilon}(s) = \epsilon_\infty \mathbf{I} + \frac{\vec{\sigma}_1(s)}{s} + \frac{\vec{\sigma}_2(s)}{s} + \dots$$

Applying the PML correction from Equation 3.31 and converting back to an equivalent conductivity, we have the following, where it is clear that the PML introduces both a free-space term and a modification to each tensor conductivity term:

$$\vec{\sigma}_{\text{pml}}(s) = s\epsilon_\infty \left(\frac{1}{\det \mathbf{S}} \mathbf{S} \mathbf{S} - \mathbf{I} \right) + \frac{1}{\det \mathbf{S}} \mathbf{S} \vec{\sigma}_1(s) \mathbf{S} + \frac{1}{\det \mathbf{S}} \mathbf{S} \vec{\sigma}_2(s) \mathbf{S} + \dots \quad (3.39)$$

A note of caution is in order, however. In anisotropic media, the PML can exhibit an instability, as noted by *Bécache et al.* [2003] and other authors [*Chevalier et al.*, 2008]. When the components of the group and phase velocities normal to the PML surface are anti-parallel, waves in the PML are exponentially amplified instead of being exponentially damped. The instability seems fundamental to the PML formulation. To the best of our knowledge, no one has yet constructed a PML that is always stable for anisotropic media.

Some recent work on absorbing boundaries has focused instead on non-local methods that attempt to directly filter the solution in k -space [*Soffer and Stucchio*, 2008]. These approaches have promise, but we do not attempt to implement them here. In cases where the instability arises, we have found an acceptable alternative for many problems is to simply break the PML by removing the gain terms from the system. This procedure converts the PML into an imperfect absorber, but it is stable and is therefore often an acceptable compromise.

3.2.4 Rotations and transformations

Our realization method for incorporating dispersive media in the DG scheme does not automatically produce the most efficient scheme possible for that medium. However, high-order methods require fewer total unknowns per degree of freedom for a given accuracy than do low-order methods. Consequently, since fewer total points are required to simulate the dispersive medium to a given accuracy, the resulting total computational cost is lower than would be possible with an optimized low-order method even though the per-point cost may increase. Nevertheless, it is still possible to exploit some prior knowledge about the medium in order to construct a scheme that requires fewer multiplications or has better sparsity patterns in the state space matrices $\{A, B, C, D\}$.

One particularly useful example is the cold plasma. A cold plasma is a gyrotropic medium, meaning that the properties of the medium are symmetric about an axis of rotation (the background magnetic field \mathbf{B}_0). If the medium is aligned such that the axis of rotation is coincident with one of the principal axes, the realization procedure generates much sparser state space matrices $\{A, B, C, D\}$ than it would if the medium were not axis-aligned.

Thus, we can realize the system in one reference frame and simply rotate it into another reference frame, reducing the total computational storage requirements. This concept is illustrated in Figure 3.6.

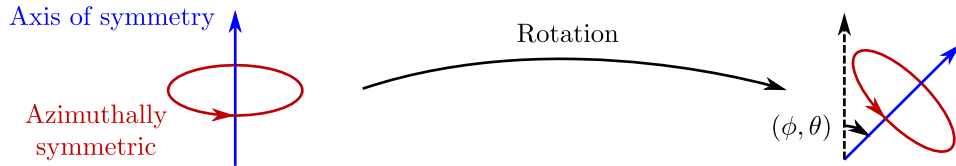


Figure 3.6: Illustration showing how a gyrotropic medium can be constructed in one reference frame (at the left) and rotated into another reference frame under only two angles (ϕ, θ) .

For instance, if we carry out the realization procedure for an axis-aligned cold

plasma, we get state space matrices of the form:

$$\mathbf{A} = \begin{bmatrix} 0 & a & 0 \\ -a & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} -b & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & b \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} -c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & c \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The matrices are not only very sparse, but only have three unique values, denoted by the constants a , b , and c . Combined with the two rotation angles ϕ and θ necessary to rotate this system into another reference frame, we have reduced the total storage requirements from 36 (one for each entry in the matrix) per point to 5 (a , b , c , ϕ , θ).

Given a realization $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ in a given reference frame and a rotation matrix \mathbf{R} that performs the rotation, the rotated realization $\{\mathbf{A}', \mathbf{B}', \mathbf{C}', \mathbf{D}'\}$ in the new reference frame is:

$$\mathbf{A}' = \mathbf{A} \tag{3.40a}$$

$$\mathbf{B}' = \mathbf{B}\mathbf{R} \tag{3.40b}$$

$$\mathbf{C}' = \mathbf{R}^T \mathbf{C} \tag{3.40c}$$

$$\mathbf{D}' = \mathbf{R}^T \mathbf{D} \mathbf{R} \tag{3.40d}$$

One other useful application of rotations and transformations is to the PML. Since the boundaries in a DG domain need not be aligned along any of the principal axes, it is useful to construct a PML to terminate some arbitrarily-oriented slab boundary. The procedure follows in the same manner. We first construct a PML in some reference frame, e.g., with surface normal $\hat{\mathbf{n}} = \hat{\mathbf{x}}$, and then rotate that canonical realization into new, rotated frames following Equation 3.40. We apply this technique to terminating curved, concave domains in Section 3.6.

3.3 Spectral properties of the DG scheme for a cold plasma

The DG scheme is *semi-discrete*, which means that it converts the spatial component of a PDE over continuous space into some operation on a finite set of unknowns on a

discrete space and does so without any coupling into the time derivative term. That is, after DG discretization, we are left with a simple ODE of the form:

$$\frac{d\mathbf{u}}{dt} = f(\mathbf{u}) \quad (3.41)$$

Equation 3.41 can then be solved using any standard time-integration method for initial value problems. Many such techniques exist, including Runge-Kutta, multistep, backwards differentiation formula (BDF), extrapolation, and the leapfrog method.

Very broadly speaking, we can divide these methods into two classes: explicit and implicit. Implicit methods incorporate information about the next timestep when computing the next timestep, that is, typically f or some function of f must be inverted. Explicit methods, by contrast, only incorporate information from the past when computing the next timestep, and thus no inversion is required. Explicit methods, however, always suffer from a stability criterion that limits the maximum timestep that can be taken in each step. The limit is essentially a sampling limit, that is, if we have a mode of the system with frequency ω , we cannot expect to reasonably sample such a mode if we take timesteps much larger than the characteristic period $1/\omega$, but with an explicit method the consequence is worse than simple inaccuracy. In fact, the system may be unstable.

In order to understand the stability criterion for our cold plasma scheme, then, we must understand what the eigenvalues of the discretized system are and what physical parameters affect them. The DG discretization, while complicated, is nonetheless linear. Thus, our DG scheme can, in principle, be written as a linear ODE of the form:

$$\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}, \quad (3.42)$$

where \mathbf{u} represents the vector of unknowns over the entire space and \mathbf{A} is some very large sparse matrix representing both the DG operator and the plasma currents. By

eigenvalue decomposition, we can consider each i^{th} mode separately:

$$\frac{d\mathbf{v}_i}{dt} = \omega_i \mathbf{v}_i \quad (3.43)$$

Note that \mathbf{v}_i denotes the i^{th} eigenvector associated with the i^{th} eigenvalue ω_i . The set of all possible ω_i is called the *eigenvalue spectrum*. We use this spectrum to determine the stability properties of a given time-stepping technique when applied to the discretized DG system.

In the absence of a plasma, the upper bound on the maximum eigenvalue is approximately equal to the ratio of the speed of light in the system and the grid spacing:

$$\omega_{\text{fs}} = \kappa \frac{c}{\min(\Delta x)},$$

where Δx is the grid spacing, c is the speed of light in the domain, and κ is some constant dependent on the order of the scheme and the location of the interpolation control points. The inverse of ω_{fs} is proportional to the maximum stable timestep when using an explicit time-stepping method.

When filled with a cold plasma, however, the situation is much more complicated. The plasma system has internal poles and zeros of its own, which can strongly distort the DG spectrum from its free-space configuration. This distortion can result in a significant reduction in the maximum allowable timestep.

To help illustrate this issue, we first plot the eigenvalue spectrum for a simple free-space case. We discretize a one meter by one meter square box into triangles using the DG scheme with a full upwinding flux ($\alpha = 1.0$ in Equation 3.18) and plot the complex eigenvalues of the resulting DG operator in Figure 3.7. Each dot on this plot shows the location of a numerical mode. Points lying exactly on the real axis correspond to oscillating solutions, while points lying on the imaginary axis correspond to evanescent solutions. In continuous space, there would be an infinite number of modes, each corresponding to a specific cavity resonance. In the discretized space, however, there are only a finite number of modes. Some of these modes map directly into physical modes, while others (particularly modes that are poorly resolved on the given grid) are nonphysical.

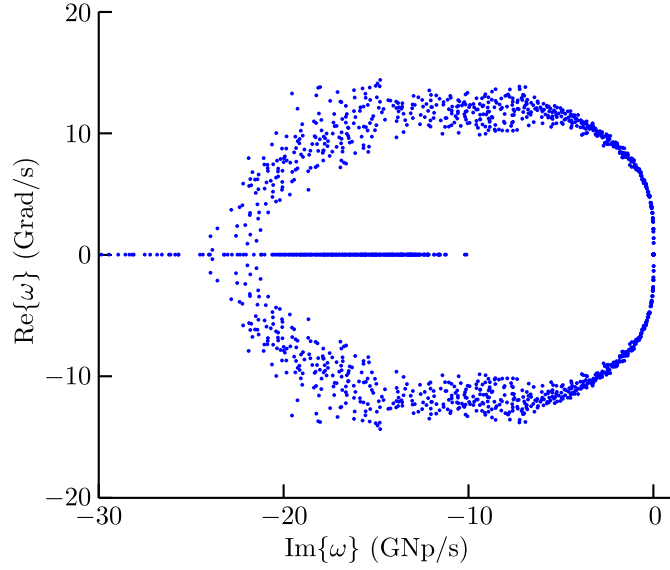


Figure 3.7: Representative DG spectrum for free-space filled 2D square domain, one meter on a side. The blue dots indicate the locations of the eigenvalues. The vertical axis corresponds to oscillatory solutions. The horizontal axis corresponds to evanescent solutions.

For a given explicit time-stepping method to be stable, all of these eigenvalues must lie entirely within a bounded *region of stability* in the complex ω -plane. The size and shape of this region is dependent on the timestep and the specific details of the time-stepping scheme being used. Two such regions of stability are shown and discussed in the next section (see Figure 3.10). For the time being, however, it suffices to use the maximum extent of the real part of the eigenvalue spectrum as a proxy for the free-space grid frequency ω_{fs} and its inverse $1/\omega_{\text{fs}}$ as a proxy for the maximum timestep. For our sample case shown in Figure 3.7, ω_{fs} is approximately 13 Grad/s.

To illustrate the consequences of including cold plasma currents, we plot two representative samples of the eigenvalue spectra for the same domain filled with a cold electron plasma in Figures 3.8 and 3.9. For reference, we show the location of some important cutoff frequencies and resonances. In addition to the plasma and cyclotron frequencies, we also show the right and left-hand cutoff frequencies, which

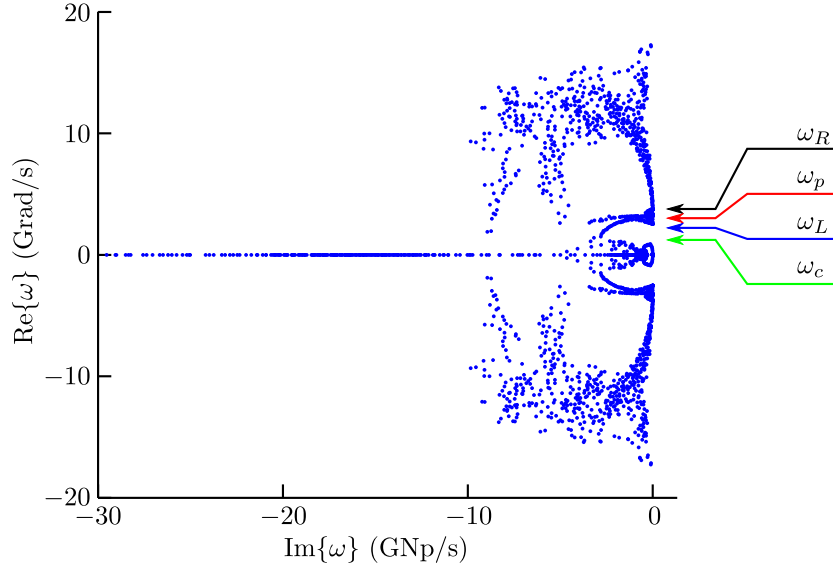


Figure 3.8: Representative DG cold plasma spectrum for a well-resolved system, where the plasma characteristic frequencies are well below ω_{fs} . The blue dots indicate the locations of the eigenvalues. The indicated frequencies ω_R , ω_p , ω_L , ω_c are the right-hand mode cutoff, plasma frequency, left-hand mode cutoff, and cyclotron frequency, respectively.

are combinations of the characteristic frequencies ω_p and ω_c :

$$\omega_R = \sqrt{\omega_p^2 + \frac{\omega_c^2}{4}} + \frac{\omega_c}{2} \quad (3.44a)$$

$$\omega_L = \sqrt{\omega_p^2 + \frac{\omega_c^2}{4}} - \frac{\omega_c}{2} \quad (3.44b)$$

Figure 3.8 illustrates the case when the free-space grid frequency ω_{fs} is much larger than these characteristic frequencies and cutoffs. While the low-frequency portion of the spectrum is highly distorted by the addition of the plasma currents, the maximum extent of the real part of the spectrum (approximately 17 Grad/s) remains relatively unchanged from its free-space value of 13 Grad/s.

In our second case (Figure 3.9), we scale both the electron number density and background magnetic field so that the characteristic frequencies of the plasma are on the order of, or larger, than the free-space grid frequency ω_{fs} . In this case, the maximum extent of the eigenvalue spectrum increases to 27 Grad/s, which has the effect of approximately halving the maximum timestep. As is apparent from Figure

3.9, it is no longer the grid spacing that dictates the maximum timestep. Instead, it is the cutoff frequency ω_R for the upper right-hand mode branch. This result can be explained by considering the behavior of the right-hand mode near cutoff. Since $k=0$ at a cutoff, a right-hand mode propagating near cutoff can be easily resolved even on an extremely coarse grid. However, while the spatial frequency of the mode is extremely low, the temporal frequency can nonetheless be quite high, which has the effect of limiting the maximum timestep.

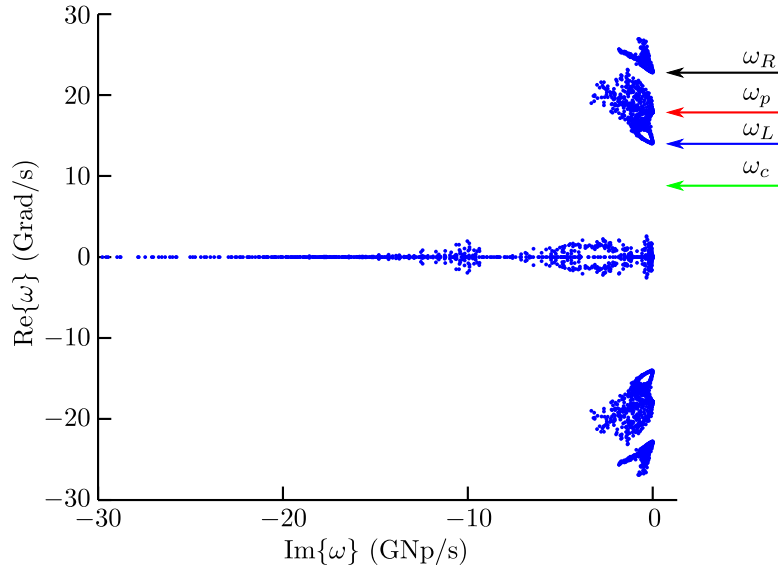


Figure 3.9: Representative DG cold plasma spectrum for a poorly-resolved system, where the characteristic frequencies are higher than ω_{fs} . The parameters plotted are the same as those plotted in Figure 3.8.

If we are interested only in resolving the low-frequency modes (as is often the case when modeling phenomena at VLF frequencies), large cutoff frequencies can impose a severe timestep restriction, increasing the simulation time by a factor of ten or more. In numerical ODE parlance, such a situation is a classic *stiff problem*, which means we are interested in low-frequency behavior, but the system imposes a timestep restriction on the order of the period of the high-frequency dynamics.

3.4 Time integration and stiffness

As demonstrated by Figure 3.9, the addition of cold plasma currents to the DG system can introduce strong stiffness if the desired frequencies of interest are much less than the plasma characteristic frequencies. Such is frequently the case for VLF signals in the ionosphere, where the frequencies of interest are in the range of 3-30 kHz, but the plasma and gyrofrequencies can be in the hundreds of kHz or even higher. For efficiency, we must deal with this issue of stiffness. In this dissertation, we are primarily concerned with systems where stiffness is imposed by the plasma characteristics.

3.4.1 Standard Runge-Kutta method

An extremely popular technique due to *Runge* [1912, p. 133], the Runge-Kutta fourth order (RK4) method, is a simple four-stage method that uses intermediate estimates of the timestep to accurately estimate the solution at some small timestep in the future. The classic RK4 method has 4 intermediate stages:

$$\begin{aligned}
 \mathbf{k}^{(1)} &= \Delta t f(\mathbf{u}^{(n)}, t^{(n)}) \\
 \mathbf{k}^{(2)} &= \Delta t f\left(\mathbf{u}^{(n)} + \frac{1}{2}\mathbf{k}^{(1)}, t^{(n)} + \frac{1}{2}\Delta t\right) \\
 \mathbf{k}^{(3)} &= \Delta t f\left(\mathbf{u}^{(n)} + \frac{1}{2}\mathbf{k}^{(2)}, t^{(n)} + \frac{1}{2}\Delta t\right) \\
 \mathbf{k}^{(4)} &= \Delta t f(\mathbf{u}^{(n)} + \mathbf{k}^{(3)}, t^{(n)} + \Delta t) \\
 \mathbf{u}^{(n+1)} &= \mathbf{u}^{(n)} + \frac{1}{6}\mathbf{k}^{(1)} + \frac{1}{3}\mathbf{k}^{(2)} + \frac{1}{3}\mathbf{k}^{(3)} + \frac{1}{6}\mathbf{k}^{(4)}
 \end{aligned} \tag{3.45}$$

The superscript (i) denotes an intermediate-stage solution that is discarded after a timestep. The superscript (n) denotes the current timestep, while $(n+1)$ denotes the solution at the next timestep (the one we wish to solve for). Since this scheme is discrete, we can assume a modal solution to analyze the stability properties, that is, we assume that the solution at a timestep n can be written in terms of a discrete-time

complex frequency Ω_i and some characteristic \mathbf{v}_i :

$$\mathbf{v}_i^{(n)} = \Omega_i^n \mathbf{v}_i^{(0)} \quad (3.46)$$

Stability is guaranteed provided all $|\Omega_i| \leq 1$. The complex frequencies Ω are found by solving for the roots of the characteristic polynomial:

$$\Omega = \frac{1}{24}(\omega\Delta t)^4 + \frac{1}{6}(\omega\Delta t)^3 + \frac{1}{2}(\omega\Delta t)^2 + (\omega\Delta t) + 1 \quad (3.47)$$

Plotted as a function of $(\omega\Delta t)$, $|\Omega| = 1$ defines the stability boundary for the scheme, which we have plotted in Figure 3.10. The eigenvalues of the cold plasma DG scheme (e.g., Figures 3.8 and 3.9) must be entirely contained within this boundary for the scheme to be stable.

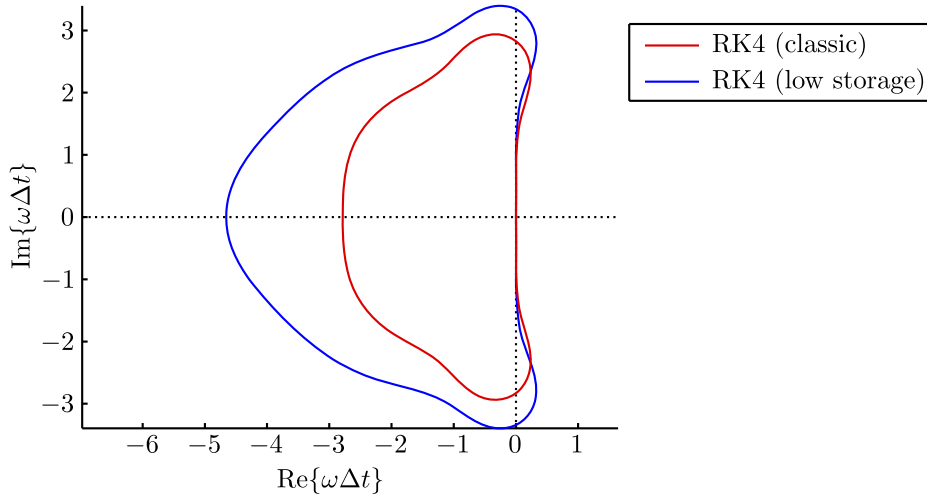


Figure 3.10: Stability boundaries for two RK4 schemes. The red boundary shows the classic RK4 stability boundary, while the blue is for the low-storage RK4 scheme.

3.4.2 Low-storage Runge-Kutta method

The classic RK4 method requires storage of all of the intermediate variables $\mathbf{k}^{(i)}$. A more memory efficient version due to *Carpenter and Kennedy* [1994] reduces the

memory requirements but at the cost of one additional function evaluation.

$$\begin{aligned}
\mathbf{p}^{(0)} &= \mathbf{u}^{(n)} \\
\mathbf{k}^{(i)} &= a_i \mathbf{k}^{(i-1)} + \Delta t f(\mathbf{p}^{(i-1)}, t^{(n)} + c_i \Delta t) & i = 1, \dots, 5 \\
\mathbf{p}^{(i)} &= \mathbf{p}^{(i-1)} + b_i \mathbf{k}^{(i)} & i = 1, \dots, 5 \\
\mathbf{u}^{(n+1)} &= \mathbf{p}^{(5)}
\end{aligned} \tag{3.48}$$

The coefficients are given in Appendix C, Section C.1. The characteristic mode Ω for this reduced-storage method is:

$$\Omega = \frac{1}{200}(\omega \Delta t)^5 + \frac{1}{24}(\omega \Delta t)^4 + \frac{1}{6}(\omega \Delta t)^3 + \frac{1}{2}(\omega \Delta t)^2 + (\omega \Delta t) + 1 \tag{3.49}$$

The boundary of stability of the scheme (where $|\Omega| = 1$) is also shown in Figure 3.10. The low-storage RK4 method has a larger region of stability on the real axis, which is an advantage for the upwind DG scheme. As illustrated by Figures 3.8 and 3.9, the upwind DG method has a large number of very large eigenvalues lying directly on the real axis, which could adversely impact the maximum timestep of the classic RK4 method.

3.4.3 IMEX Runge-Kutta method

The basic Runge-Kutta approach can be extended to split schemes, one useful example of which is a singly diagonally implicit IMEX (IMplicit-EXplicit) scheme. The IMEX-RK method can be used to solve certain types of stiff problems in an extremely efficient manner [Kanevsky *et al.*, 2007; Kennedy and Carpenter, 2003]. The scheme is used to solve ODEs of the following form, where the right-hand side (RHS) has been split into the sum of two parts:

$$\frac{d\mathbf{u}}{dt} = f^{[\text{im}]}(\mathbf{u}) + f^{[\text{ex}]}(\mathbf{u}) \tag{3.50}$$

The scheme steps one part of the RHS $f^{[\text{ex}]}$ using an explicit method while the other part $f^{[\text{im}]}$ is solved implicitly. The errors are cancelled using the standard Runge-Kutta approach so the overall scheme retains its high order accuracy. The implicit

part has no timestep restriction, so if the function $f^{[\text{im}]}$ can be easily inverted, the overall scheme can be highly efficient even for stiff problems, provided a suitable splitting can be found.

The scheme is as follows, where the different sets of coefficients are denoted with the superscripts $[\text{im}]$ and $[\text{ex}]$:

$$\begin{aligned} \mathbf{u}^{(i)} = \mathbf{u}^{(n)} + \Delta t \sum_{j=1}^s a_{ij}^{[\text{ex}]} f^{[\text{ex}]}(t^{(n)} + c_j \Delta t, \mathbf{u}^{(j)}) + \\ \Delta t \sum_{j=1}^s a_{ij}^{[\text{im}]} f^{[\text{im}]}(t^{(n)} + c_j \Delta t, \mathbf{u}^{(j)}) \end{aligned} \quad (3.51a)$$

$$\begin{aligned} \mathbf{u}^{(n+1)} = \mathbf{u}^{(n)} + \Delta t \sum_{j=1}^s b_j f^{[\text{ex}]}(t^{(n)} + c_j \Delta t, \mathbf{u}^{(j)}) + \\ \Delta t \sum_{j=1}^s b_j f^{[\text{im}]}(t^{(n)} + c_j \Delta t, \mathbf{u}^{(j)}) \end{aligned} \quad (3.51b)$$

The scheme we use is a six-stage ($s=6$), fourth order method with singly diagonally-implicit stages. The coefficients are given in Appendix C, Section C.2. At each RK stage for $i > 1$, we must solve an implicit system of the form:

$$\mathbf{u}^{(i)} - (\Delta t) a_{ii}^{[\text{im}]} f^{[\text{im}]}(\mathbf{u}^{(i)}) = \mathbf{r},$$

where the right-hand side \mathbf{r} collects all of the summed terms that can be evaluated explicitly. Or for a linear system, we can write the function as a matrix \mathbf{A} , so:

$$\mathbf{u}^{(i)} = \left[\mathbf{I} - (\Delta t) a_{ii}^{[\text{im}]} \mathbf{A} \right]^{-1} \mathbf{r} \quad (3.52)$$

For this IMEX scheme, all coefficients a_{ii} for $i > 1$ equal exactly $1/4$, so we can re-use the matrix inverse or factorization at each stage.

3.4.4 Cold plasma IMEX method

Our approach for combating plasma-induced stiffness is to use the IMEX method to integrate the plasma currents implicitly while keeping the DG part explicit. The

splitting we use, illustrated for a single species, is:

$$\frac{d}{dt} \begin{bmatrix} \mathbf{x} \\ \mathbf{E} \\ \mathbf{H} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{B} & 0 \\ -\frac{1}{\epsilon}\mathbf{C} & -\frac{1}{\epsilon}\mathbf{D} & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\text{Implicit part}} \begin{bmatrix} \mathbf{x} \\ \mathbf{E} \\ \mathbf{H} \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\epsilon}\nabla \times \\ 0 & -\frac{1}{\mu}\nabla \times & 0 \end{bmatrix}}_{\text{Explicit part}} \begin{bmatrix} \mathbf{x} \\ \mathbf{E} \\ \mathbf{H} \end{bmatrix}$$

From Equation 3.52, at each RK stage $i > 1$, we must solve a system of the following form, where for completeness we now consider a multi-species plasma with the realization for the k^{th} species characterized by the four matrices $\{\mathbf{A}_k, \mathbf{B}_k, \mathbf{C}_k, \mathbf{D}_k\}$ and state vector \mathbf{x}_k :

$$\begin{bmatrix} \mathbf{I} - (\Delta t)a_{ii}\mathbf{A}_1 & 0 & \cdots & -(\Delta t)a_{ii}\mathbf{B}_1 \\ 0 & \mathbf{I} - (\Delta t)a_{ii}\mathbf{A}_2 & \cdots & -(\Delta t)a_{ii}\mathbf{B}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\epsilon}(\Delta t)a_{ii}\mathbf{C}_1 & \frac{1}{\epsilon}(\Delta t)a_{ii}\mathbf{C}_2 & \cdots & \mathbf{I} + \frac{1}{\epsilon}(\Delta t)a_{ii}(\mathbf{D}_1 + \mathbf{D}_2 + \dots) \end{bmatrix} \begin{bmatrix} \mathbf{x}_1^{(i)} \\ \mathbf{x}_2^{(i)} \\ \vdots \\ \mathbf{E}^{(i)} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_{x1} \\ \mathbf{r}_{x2} \\ \vdots \\ \mathbf{r}_E \end{bmatrix}$$

Or simply:

$$\tilde{\mathbf{M}}\mathbf{u} = \mathbf{r} \quad (3.53)$$

The vector \mathbf{r} collects all of the right-hand side terms while \mathbf{u} collects all of the unknowns. $\tilde{\mathbf{M}}$ is a large block matrix. Recalling that \mathbf{I} is the identity matrix, this matrix has all diagonals filled and is invertible, with condition number dependent on the timestep. This matrix has a block arrow sparsity pattern and can thus be inverted efficiently using a block inverse. For convenience, we make the following definitions:

$$\begin{aligned} \tilde{\mathbf{A}}_k &= \mathbf{I} - (\Delta t)a_{ii}\mathbf{A}_k \\ \tilde{\mathbf{B}}_k &= -(\Delta t)a_{ii}\mathbf{B}_k \\ \tilde{\mathbf{C}}_k &= \frac{1}{\epsilon}(\Delta t)a_{ii}\mathbf{C}_k \\ \tilde{\mathbf{D}} &= \mathbf{I} + \frac{1}{\epsilon}(\Delta t)a_{ii}(\mathbf{D}_1 + \mathbf{D}_2 + \dots) \end{aligned}$$

The Schur complement of the system is:

$$\tilde{S} = \tilde{D} - \tilde{C}_1 \tilde{A}_1^{-1} \tilde{B}_1 - \tilde{C}_2 \tilde{A}_2^{-1} \tilde{B}_2 - \dots$$

Defining an intermediate solution \mathbf{b} :

$$\mathbf{b} = \mathbf{r}_E - \tilde{C}_1 \tilde{A}_1^{-1} \mathbf{r}_{x1} - \tilde{C}_2 \tilde{A}_2^{-1} \mathbf{r}_{x2} - \dots$$

The electric field at stage i is then found by:

$$\mathbf{E}^{(i)} = \tilde{S}^{-1} \mathbf{b}$$

We can then solve for the state vectors $\mathbf{x}_k^{(i)}$ as:

$$\begin{aligned} \mathbf{x}_1^{(i)} &= \tilde{A}_1^{-1} \left(\mathbf{r}_{x1} - \tilde{B}_1 \mathbf{E}^{(i)} \right) \\ \mathbf{x}_2^{(i)} &= \tilde{A}_2^{-1} \left(\mathbf{r}_{x2} - \tilde{B}_2 \mathbf{E}^{(i)} \right) \\ &\dots \end{aligned}$$

If Δt is constant, we can pre-compute all of the matrices needed to compute the matrix inverse in Equation 3.53. Specifically, we pre-compute and store the matrices \tilde{A}_k^{-1} , $\tilde{A}_k^{-1} \tilde{B}_k$, and $\tilde{C}_k \tilde{A}_k^{-1}$ per species per point and the matrix \tilde{S}^{-1} per point, roughly doubling the material coefficient storage requirements.

This IMEX time integration technique is more expensive per timestep than the low-storage RK method discussed in the previous section because the IMEX method has more stages and each stage involves more matrix multiplications and sums. However, it permits taking timesteps as large as allowed by the DG grid, which can result in significant speedups for very stiff plasmas.

3.4.5 Time integration convergence tests

To demonstrate fourth-order convergence of the time-stepping schemes, we discretize a plasma-filled metallic cavity and initialize it with a sinusoidal perturbation with vanishing tangential electric field at the wall, i.e., a fundamental cavity mode if the

domain were not filled with a plasma. We then compute a reference simulation on this domain using an extremely small timestep $\Delta t = 0.001\Delta t_{\max}$ up to a final time t_f . We repeat the simulation for different timesteps up to the maximum $\Delta t = \Delta t_{\max}$ and compute the norm error over the space. The results are plotted in Figure 3.11.

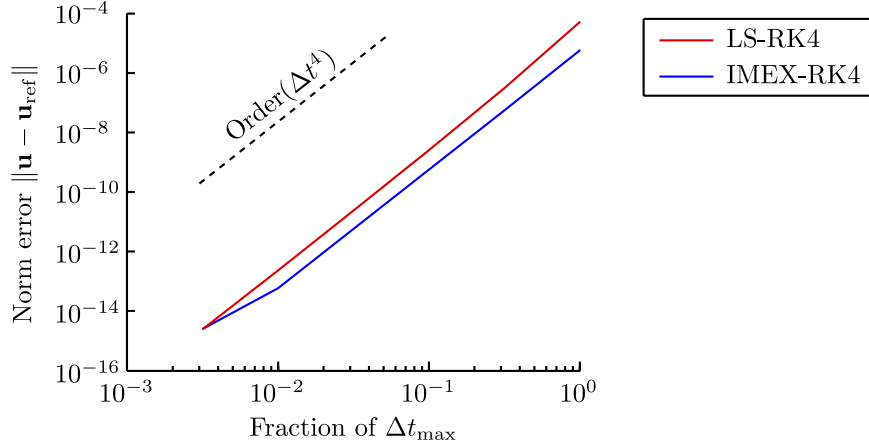


Figure 3.11: Plot demonstrating convergence of the IMEX and explicit RK4 methods for a cold electron plasma-filled domain. The horizontal axis is the fractional timestep with respect to the maximum Δt_{\max} . The vertical axis is the normed error between the reference solution and the solution run at a given fractional timestep. The deviation from order Δt^4 at low fractions is due to finite precision effects.

For reference, we show the slope for theoretical fourth-order convergence, demonstrating the convergence of both schemes. The IMEX method not only converges but also has slightly lower absolute error than the explicit method.

3.5 Plasma verification

In Section 3.4.5, we demonstrated convergence of the time integration scheme for a cold plasma. To demonstrate convergence of the scheme as a whole, we show that the error is systematically reduced as a function of the grid spacing and that this rate of convergence is higher for increasing polynomial order. As in the previous section, we compute a reference simulation on a plasma-filled domain to high order. We then compare this reference to lower-order simulations, varying both the grid spacing and the polynomial order. We show 1D results in Figure 3.12. The optimum convergence rates are nearly exactly met for the upwind scheme ($\alpha = 1.0$ in Equation

3.18). The central scheme ($\alpha=0.0$) has an apparent loss of optimal convergence for odd polynomial orders N_p , which is due to the odd-even structure of Maxwell's equations.

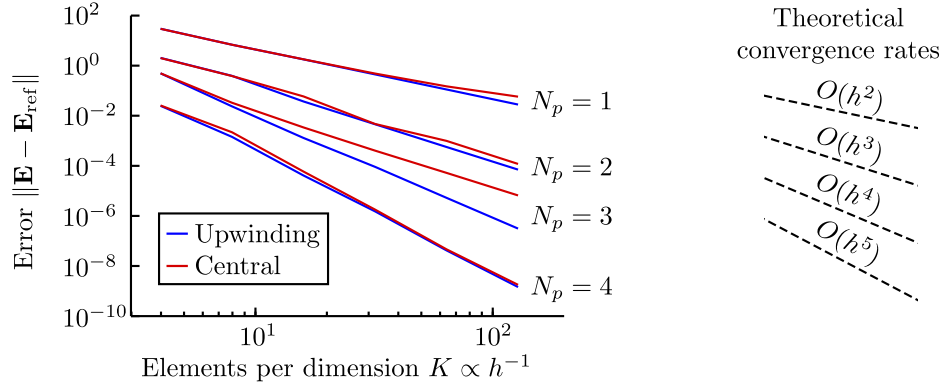


Figure 3.12: Plot demonstrating convergence of the DG cold plasma scheme. We fill a 1D domain with a cold magnetized electron plasma with background magnetic field \mathbf{B}_0 aligned along the free spatial dimension, and compute a reference simulation to high order. We then vary both the order and grid spacing h for varying polynomial orders N_p . At the left, we plot the measured error as a function of the grid spacing h for the upwind flux and central flux. At the right we show the theoretical optimum convergence rates for these polynomial orders, $O[h^{(N_p+1)}]$.

We show the same convergence test in two dimensions in Figure 3.13. The convergence results are similar to the 1D results in Figure 3.12, with the central flux in this case apparently performing nearly optimally. This difference between the 1D and 2D results is due partially to the loss of hyperbolicity (that is, some modes become evanescent) for some propagation directions in a cold magnetized plasma. For these directions, there is no actual propagating wave and the central flux becomes the optimal choice. The balance between the two tends to mask the loss of optimal convergence for the propagating modes.

The convergence results in three dimensions are so close to these cases that we omit them for brevity. In order to understand the full spectral characteristics of the scheme, we also compute a simulated dispersion diagram, plotting the allowed frequencies ω as a function of the wavenumber k . To compute this diagram, we drive the system with a narrow (in both time and space) current source at time $t=0$ and position $x=0$. We run the simulation for some time and record the resulting fields over the space for each timestep. Given this data in t - x space, we can compute the dispersion

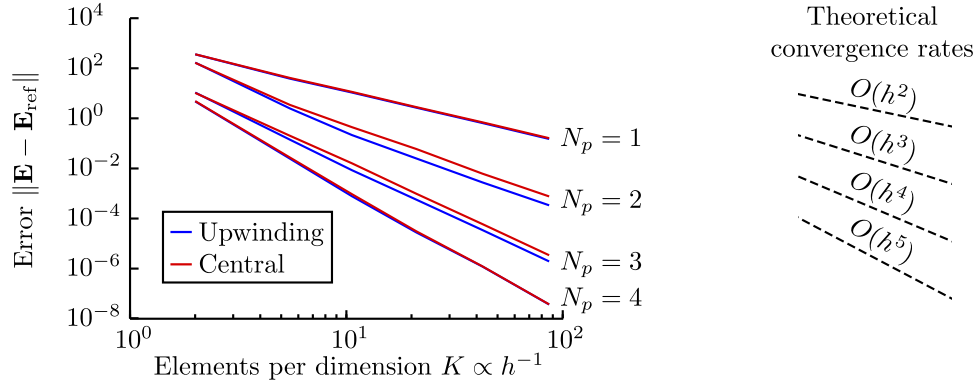


Figure 3.13: Plot demonstrating convergence of the 2D DG cold plasma scheme. We fill a 2D domain with a cold electron plasma and compute a reference simulation to high order. We then vary both the order and per-dimension grid spacing h for varying polynomial orders N_p . At the left, we plot the measured error as a function of the per-dimension grid spacing h for the upwind flux and central flux. At the right we show the theoretical optimum convergence rates for these polynomial orders, $O[h^{(N_p+1)}]$.

diagram by a 2D Fourier transform into ω - k space. This method is particularly useful because unlike the spectra in Figures 3.8 and 3.9, the synthetic dispersion diagram illuminates the full spectral characteristics of the scheme, including the relationship between the temporal and spatial parts.

For illustration purposes, we consider a system with electron number density $N = 3.2 \times 10^{10} \text{ m}^{-3}$ and background magnetic field $B_0 = 4.34 \times 10^{-5} \text{ T}$. The theoretical dispersion relation is shown in Figure 3.14 for propagation parallel and perpendicular to the background magnetic field.

In Figure 3.15, we display the computed numerical dispersion relation for the central flux method ($\alpha = 0.0$ in Equation 3.18). The numerical dispersion relation is noticeably different than that in Figure 3.14, not only diverging significantly from reality at high spatial frequencies, but also showing spurious modes, which are modes that are not present in the original equations but nonetheless appear in the numerical scheme. This phenomenon is a well-known consequence of using the central flux. Its consequences for the cold plasma scheme seem especially severe, since there exists no clear separation between all of the spurious modes and all of the natural modes.

However, the physical and spurious branches do not touch, so there is apparently no way for the physical and spurious modes to couple into each other. Provided, then,

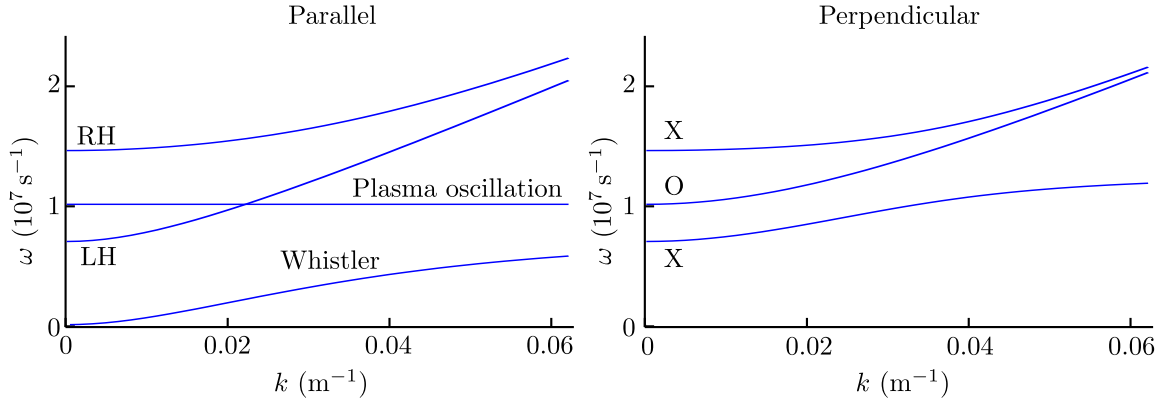


Figure 3.14: Theoretical dispersion relation for a magnetized plasma with electron number density $N = 3.2 \times 10^{10} \text{ m}^{-3}$ and background magnetic field $B_0 = 4.34 \times 10^{-5} \text{ T}$. The left plot is for propagation parallel to the background magnetic field, while the right is for propagation perpendicular. The different modes are labeled with their common names: O (ordinary mode), X (extraordinary mode), whistler, RH (right-hand mode), and LH (left-hand mode).

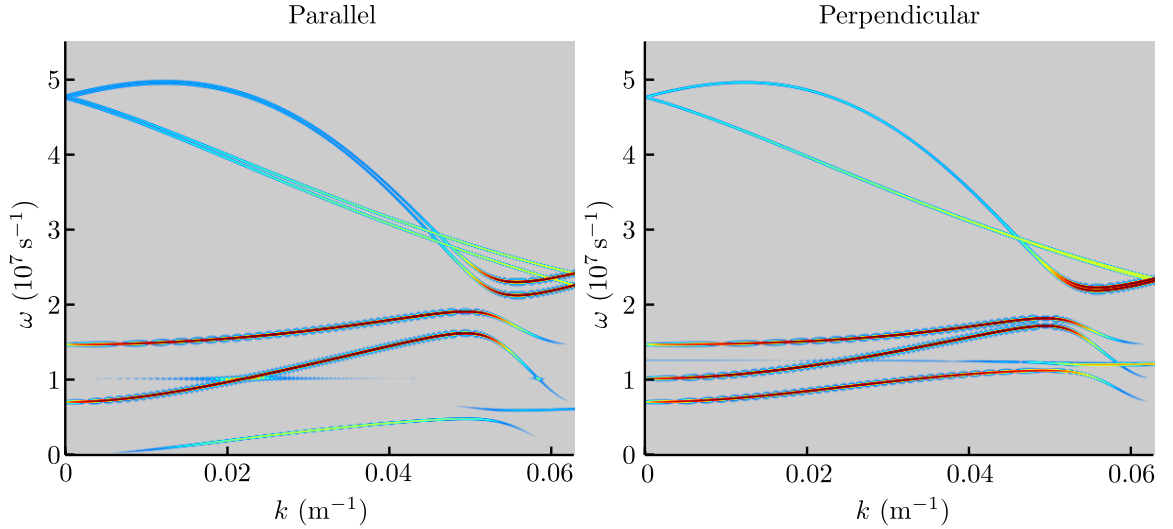


Figure 3.15: Numerical dispersion relation for a magnetized plasma using the DG cold plasma scheme with a purely central flux. The left plot is for propagation parallel to the background magnetic field, while the right is for propagation perpendicular to it. The parameters are identical to those used in Figure 3.14.

that we avoid exciting the spurious modes by other means, the physical solution can still be recovered. In this instance, we have deliberately driven the nonphysical modes by using a very poorly resolved, narrow source in time and space so that all modes, spurious and physical, are excited. If the initial conditions and sources, however, are kept smooth, the spurious modes are not likely to be excited except through

numerical roundoff error. Further, supposing there is only one mode of interest in a particular simulation, say, the whistler mode, we can reasonably filter the solution in post-processing to completely suppress the spurious mode, since the spurious branch of the whistler mode is confined to temporal frequencies strictly greater than those of the physical branch. Interestingly, the perpendicular dispersion relation shows evidence of a hybrid resonance, visible as a horizontal line in the rightmost plot in Figure 3.15 at the upper hybrid frequency $\omega_{uh} = (\omega_p^2 + \omega_c^2)^{1/2}$. This mode is physical but is not normally visible in the purely cold plasma approximation as it requires a non-zero divergence of the electric field, that is, $\nabla \cdot \mathbf{E} \neq 0$. The mode is instead driven by numerical errors at high values of k . The upwind flux discussed next cleanly suppresses this upper hybrid mode, while still preserving the plasma resonance at ω_p for parallel propagation.

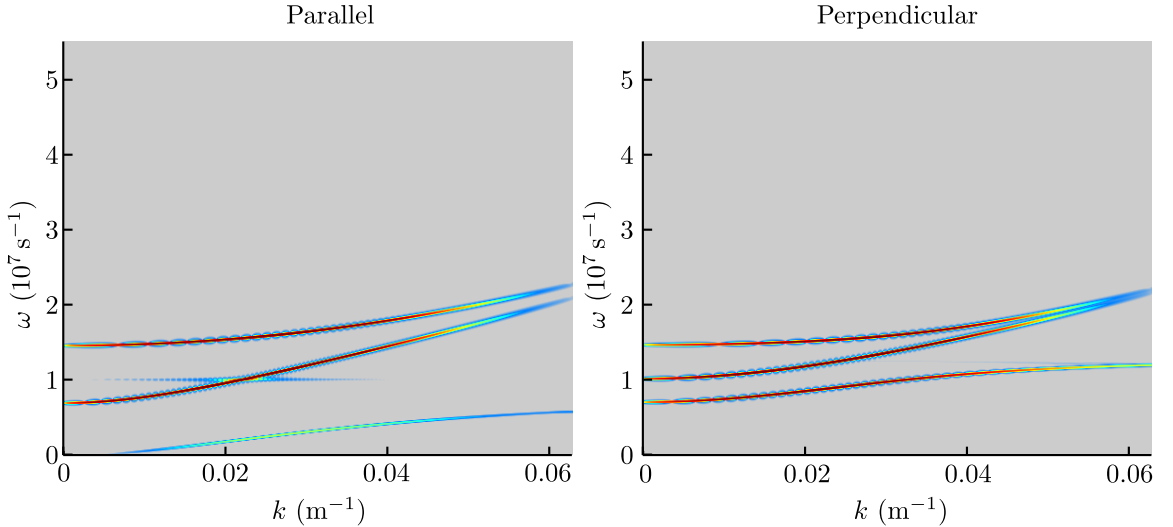


Figure 3.16: Numerical dispersion relation for a magnetized plasma using the DG cold plasma scheme with a purely upwind flux. The left plot is for propagation parallel to the background magnetic field, while the right is for propagation perpendicular to it. The parameters are identical to those used in Figure 3.14.

The results for the upwind flux ($\alpha = 1.0$ in Equation 3.18) are shown in Figure 3.16. The results are apparently much better, showing neither nonphysical spurious modes nor branches with negative group velocities. For these reasons, we typically prefer the pure upwind flux. The pure upwind flux, however, is mildly lossy, which can introduce errors for extremely sensitive scattering calculations. The central flux, by

contrast, is strictly energy-preserving. We note, however, that we have not found the upwind flux to introduce noticeable losses for any reasonably well-resolved problems at moderate order, say 10 unknowns per wavelength per dimension with a fourth-order basis.

3.6 PML verification

To further demonstrate the validity of our scheme, we test the performance of the PML implemented as described in Section 3.2.3. We perform a standard convergence test, showing that the reflection error of the PML is reduced exponentially as a function of the polynomial order (equivalently, as the number of points per dimension is increased). We compute two simulations, one on a large domain truncated with a perfect reflector, which serves as the reference solution for a second simulation on a smaller, PML-truncated grid. We place a current source at the center of the domain and run the simulation until any reflections have completely filled the smaller computational domain. The normalized reflection error $20 \log_{10}(\|\mathbf{u} - \mathbf{u}_{\text{ref}}\|/\|\mathbf{u}_{\text{ref}}\|)$ gives a global measure of the performance of the scheme.

We first demonstrate the scheme on a 2D rectangular domain. We truncate each face with a PML oriented to the surface normal. For example, for faces with surface normal $\hat{\mathbf{n}} = \hat{\mathbf{x}}$, we set the PML loss parameters (defined in Equation 3.35) $\xi_x = \xi$ and $\xi_y = \xi_z = 0$. Similarly, at faces with surface normal in the $\hat{\mathbf{y}}$ direction, we set $\xi_y = \xi$ and $\xi_x = \xi_z = 0$. In corner regions, we set $\xi_x = \xi_y = \xi$ and $\xi_z = 0$. We then compute the residual error as a function of increasing ξ and repeat for different polynomial orders N_p . The results are plotted in Figure 3.17.

The scheme shows clear convergence, with the reflection error decreasing exponentially as a function of both the loss parameter ξ and the polynomial order N_p . At low ξ , the reflection error is large because the wave is not sufficiently damped as it makes a traverse through the PML, reflects from the outer boundary, then propagates again through the PML back into the main domain. For increasing ξ , then, the damping rate increases, so the reflection error in turn decreases exponentially until saturation is reached. At saturation, the accuracy of the scheme begins to dominate the total error and increasing ξ further has no advantage and may even have a detrimental

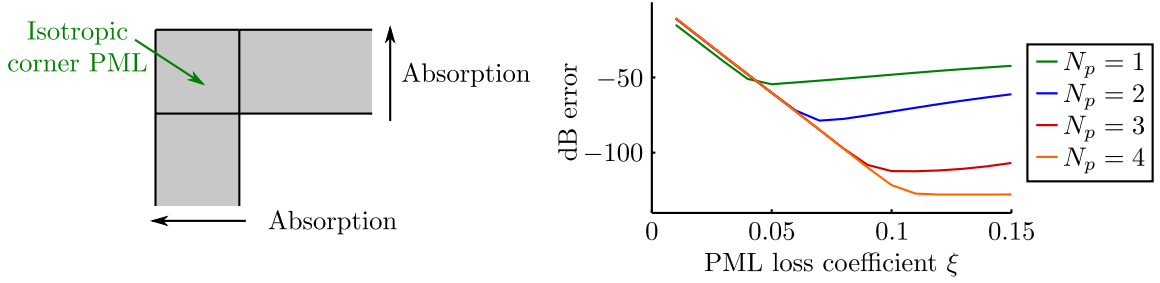


Figure 3.17: PML convergence test for a rectangular domain. **Left:** Illustration of one section of the PML-truncated domain, showing three distinct PML regions. In regions with face normal $\hat{\mathbf{n}} = \hat{\mathbf{x}}$, we set the PML loss parameter $\boldsymbol{\xi} = (\xi, 0, 0)$, leading to attenuation in the $\hat{\mathbf{x}}$ direction. For face normal $\hat{\mathbf{n}} = \hat{\mathbf{y}}$, we set $\boldsymbol{\xi} = (0, \xi, 0)$. In the corner regions, we set $\boldsymbol{\xi} = (\xi, \xi, 0)$. **Right:** Plot of the normalized error for the PML-truncated rectangular domain, shown in dB as a function of increasing PML loss parameter ξ for varying polynomial order N_p .

effect on the overall reflection error. As a function of increasing polynomial order N_p (roughly proportional to the number of points per dimension), the minimum reflection error attained decreases exponentially, i.e., the scheme converges. Increasing the polynomial order beyond four does not lead to any additional improvement for this particular mesh, because the reflection error is already so low that finite precision effects begin to dominate.

We further note that we construct the PML with no grading, as is often done in FDTD simulations. The PML is encountered as a single, abrupt jump in the material parameters. Nevertheless, the scheme still performs exceptionally well, with reflection errors at least as low as for graded FDTD schemes, demonstrating the better handling of material discontinuities by the use of the DG scheme.

As a further demonstration of the flexibility of the direct realization approach, we use our method to construct a PML for an arbitrary, curved concave boundary. Our test domain is illustrated in Figure 3.18, where we show the electric fields produced by a monochromatic current source in both an untruncated curved domain and a PML-truncated curved domain.

We first construct a reference PML to truncate an $\hat{\mathbf{x}}$ -oriented surface. Using Equation 3.40, we rotate this reference PML such that the surface normal is oriented normally to some arbitrarily-rotated slab boundary. In the corner regions, we use an isotropic corner absorber as described earlier, with $\boldsymbol{\xi} = (\xi, \xi, 0)$. This corner absorber

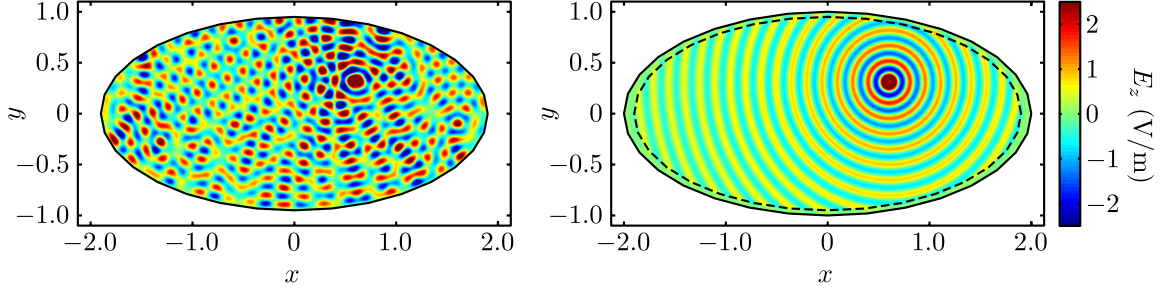


Figure 3.18: Illustration of a PML truncating a curved domain. We place a vertical ($\hat{\mathbf{z}}$ -oriented) current source in an ellipse-shaped domain and plot the resulting vertical electric fields after sufficient time has passed to fill the domain with reflections. **Left:** A curved domain with a naive (perfectly reflecting) truncation, showing strong reflection errors. The dark oval line shows the location of the boundary. **Right:** Curved domain truncated with a PML. The dark line shows the location of the reflective boundary, while the dashed line shows the PML interface. Virtually no reflection errors are visible.

is invariant under rotations about the $\hat{\mathbf{z}}$ axis, so the corner regions all have identical material properties. The technique and convergence results are illustrated in Figure 3.19.

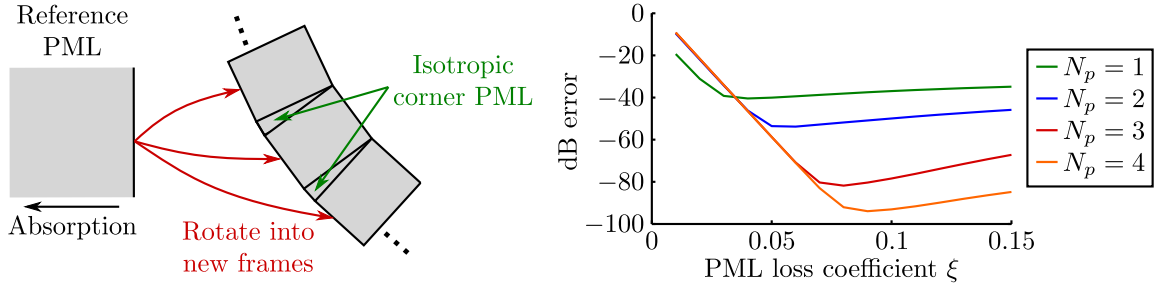


Figure 3.19: PML convergence test for a curved domain. **Left:** Illustration of the faceted PML. We begin with a reference PML aligned along the $\hat{\mathbf{x}}$ direction. We construct each face slab by rotating this reference PML into a new frame such that the normal coincides with that slab's face normal. The corner regions are truncated as before, with PML loss parameter $\boldsymbol{\xi} = (\xi, \xi, 0)$. **Right:** Plot of the normalized error for the PML-truncated rectangular domain, shown in dB as a function of increasing PML loss parameter ξ for varying polynomial order N_p .

3.7 Implementation notes

The efficiency of a scheme depends not only on its formulation but on its implementation as well, and therefore a short discussion on implementation is in order.

Each Runge-Kutta stage involves computation of a difference between the fields

on the faces of adjoining elements, $\mathbf{d}^- = \mathbf{u}^- - \mathbf{u}^+$, or in index notation:

$$d_{i-} = u_{i-} - u_{i+} \quad (3.54)$$

The index set \mathbf{i}^- denotes face indices interior to each element, while \mathbf{i}^+ denotes those exterior (facing) each element. The face indices and their matching exterior neighbors are built in a pre-processing step. The difference operation is performed using a generic vector scatter using the Petsc library [Balay *et al.*, 2011, 2010, 1997], which performs the following operation on arbitrary vectors \mathbf{a} and \mathbf{b} :

$$a_j = b_k \quad (3.55)$$

The index sets \mathbf{j} and \mathbf{k} are arbitrary sets of indices with conforming sizes. The amount of communication between nodes must also be minimized for efficient parallelization, as illustrated in Figure 3.20. We pre-segment the domain with the ParMETIS graph

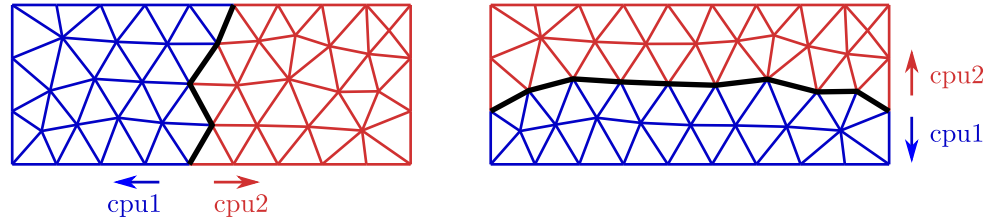


Figure 3.20: **Left:** An efficient segmentation of a grid onto two CPUs. Only four faces cross the CPU boundary. **Right:** An inefficient segmentation, where nine faces cross the CPU boundary.

partitioning library [Karypis *et al.*, 1997], which approximately solves the weighted graph partitioning problem in parallel.

Each element-local operation requires multiplying a small matrix (e.g., \mathbf{M}_1) and a small number of local field values. We accelerate this operation by concatenating the unknowns into another large, dense matrix \mathbf{M}_2 , as shown in Figure 3.21. Clever blocking strategies can improve the speed of this dense matrix operation by many times over naive implementations. We use the ATLAS library [Whaley and Petitet, 2005; Whaley *et al.*, 2001; Whaley and Dongarra, 1999, 1998, 1997], an automatically-tuned linear algebra package with optimized dense matrix-matrix multiplication routines.

We also use similar tuning strategies to accelerate more complicated operations.

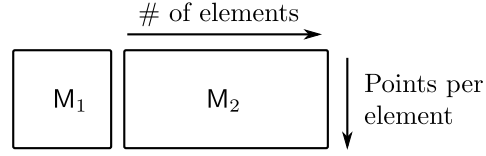


Figure 3.21: Illustration of the local DG matrix vector operations converted into a single dense matrix multiplication.

For instance, we may have a function composition representing two parts of the update step for a vector of unknowns \mathbf{u} :

$$\mathbf{u} \leftarrow f(g(\mathbf{u}))$$

It is not always clear whether it is more efficient to first compute an intermediate vector from the function g , then use that to compute the update to \mathbf{u} , i.e.,

$$\begin{aligned} \mathbf{g} &\leftarrow g(\mathbf{u}) \\ \mathbf{u} &\leftarrow f(\mathbf{g}), \end{aligned}$$

or to compute the intermediate quantity as a simple scalar within a loop, i.e., for each index i , to compute:

$$\begin{aligned} g_i &\leftarrow g(u_i) \\ u_i &\leftarrow f(g_i) \end{aligned}$$

The answer, somewhat unsurprisingly, is that it depends greatly on the size of the local working set and how much shared data is used by the functions g and f . Since the CPU cache is many times faster than main memory, it is more efficient to perform a complete set of operations on an active working set in cache before moving onto a new block of data, which requires another fetch from main memory.

We have found that a configurable blocking strategy leads to the most efficient evaluation of each stage, where the size of the intermediate vector is adjusted empirically so that cache utilization is optimized. We show the efficiency of this strategy in Figure 3.22 for a sample problem in two dimensions. For a given fixed grid, we vary

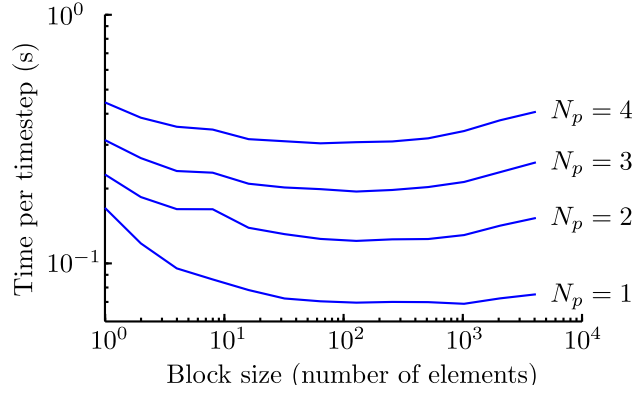


Figure 3.22: Total CPU time per timestep plotted as a function of the local working set block size in log-log scale. We plot the time per timestep for varying polynomial orders N_p .

the local block size, run a simulation for some number of timesteps, and then plot the real CPU time per timestep. We repeat this procedure for varying polynomial orders from $N_p = 1$ to $N_p = 4$.

The optimum for Maxwell's equations is attained somewhere near block sizes of approximately 100 for $N_p = 4$, with a slight increase in the optimum as the polynomial order decreases due to the decrease in size of the mass and stiffness matrices. At small block sizes, we are not efficiently taking advantage of the ATLAS library's optimized matrix-matrix products, so the scheme suffers. At large block sizes, the working set begins to exceed the CPU cache size and the scheme once again suffers. The total performance gains are relatively modest when the scheme is considered as a whole, since the element-local operations are only part of the entire computational cost (the other major contributor being inter-element communication). However, speedup factors of two or greater are nonetheless possible.

Finally, we plot the speedup of our scheme as a function of the number of CPUs, demonstrating that the scheme is efficiently parallelized. Results for a typical mesh are shown in Figure 3.23. The speedup is approximately linear over a very wide range of local problem sizes, only becoming sublinear when the number of elements per CPU is lower than approximately 400 (the red line in Figure 3.23). At this point, the communication cost per timestep begins to dominate the overall computational cost and we reach a point of diminishing returns.

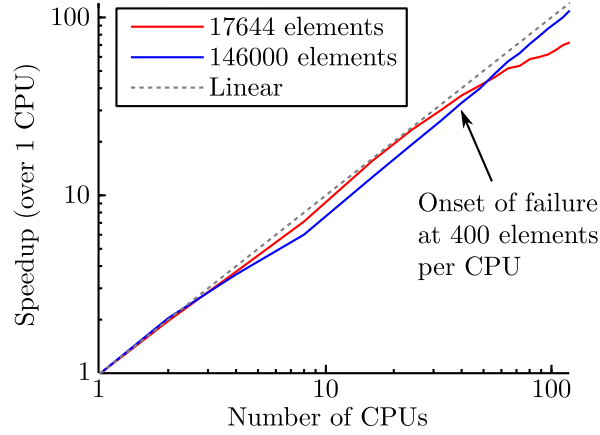


Figure 3.23: Speedup (referenced to one CPU) vs. number of CPUs for a typical DG domain. The optimum case is a perfect linear speedup, that is, the scheme is faster in direct proportion to the number of CPUs used. This optimum is indicated with a dashed line. The plots above indicate that the speedup is approximately linear down to approximately 400 elements per CPU. These timing tests are conducted on a small 16-node cluster connected via an Infiniband interconnect.

3.8 Summary

In this chapter, we have described the application of the nodal DG technique [Hesthaven and Warburton, 2002] to Maxwell’s equations in anisotropic, dispersive media. We have described our new auxiliary differential equation method to incorporate any linear material in the DG framework. We have demonstrated the validity of the scheme as applied to two types of linear materials: the perfectly matched layer (PML) and a cold, multi-component magnetized plasma. Since all fields are co-located and all material properties are incorporated as a set of auxiliary differential equations, the resulting scheme is semi-discrete and completely explicit.

Our approach allows us to directly address the problem of plasma stiffness. If the frequencies of interest are far below the characteristic frequencies of the material, the maximum timestep allowed by an explicit method is limited by these characteristic frequencies, which would normally severely limit the range of problems that we could efficiently solve. To illustrate this issue and its consequences for simulating propagation in cold plasmas, we have explicitly computed the eigenvalue spectra for some representative examples, demonstrating that the cold plasma currents introduce eigenvalues with magnitude on the order of the cutoff frequency ω_R for the

upper right-hand mode.

In our framework, we address this stiffness problem by simply replacing the classical RK4 time-stepping scheme with an implicit-explicit (IMEX) RK4 scheme. This scheme allows us to circumvent the timestep restriction without the need to resort to extremely inaccurate and diffusive operator-splitting approaches (e.g., Strang splitting). We have further described a block decomposition approach to pre-computing the matrix inverses needed by the IMEX scheme, which leads to only a doubling of the material coefficient storage requirements and only modest increases in the total computational cost per timestep.

We have illustrated the convergence of our method in both time and space for plasma-filled domains, showing that the schemes converge at their theoretically optimum rates. Finally, through the use of numerical dispersion relations, we have shown that the schemes are consistent with cold plasma theory. When central fluxes are used, the numerical dispersion relations show evidence of complicated spurious modes. The central DG scheme in free space also shows a spurious mode, but since there is only one free-space mode and one spurious mode, the spurious mode is easily detected and even filtered out of the solution. For our plasma scheme, one additional spurious mode appears for each real plasma mode. The upwind scheme does not suffer from this problem, reproducing only the physical cold plasma modes. However, the upwind scheme suffers from artificial numerical losses at high (poorly-resolved) spatial frequencies.

As an additional application, we have also demonstrated a PML implementation. The scheme converges well, with normalized reflection errors lower than -100 dB. Our approach also allows us to construct a novel conformal PML, which allows the termination of arbitrarily-shaped concave domains with no modification to the basic scheme.

Finally, we have concluded this chapter with a brief discussion of some implementation details, with particular attention paid to parallelization.

Chapter 4

The DG-PIC method

Many interesting problems in plasma physics are extremely difficult to solve either analytically or numerically. From a numerical standpoint, the fundamental difficulty is *dimensionality*. As discussed in Section 2.1.3, in principle all behavior of a collisionless plasma is completely described by the Vlasov equation (Equation 2.12) coupled with Maxwell's equations. Unfortunately, the Vlasov equation describes the time evolution of an unknown phase space density in six dimensions: three spatial and three momentum. At the present time, direct discretization of this equation is intractable for large problems since the memory and computational costs scale as N^6 , where N is the number of discretized unknowns per dimension. As such, a number of approximate techniques are of necessity. These fall into four general classes of techniques: direct discretization, linearization, multi-moment expansions, and particle techniques.

The direct method involves direct discretization of the Vlasov equation in either its conservative form (Equation 2.12) or its non-conservative form (Equation 2.13). Since a naive discretization requires on the order of N^6 unknowns, this approach is only feasible if either the dimensionality can be reduced, e.g., by reducing either the number of spatial dimensions, or if the solution space itself is extremely sparse, which, when combined with adaptive techniques, leads to less expensive schemes.

The linearized Vlasov equation (Equation 2.14) is a powerful analytical tool for determining the small-signal behavior of a plasma but is not often used directly as a simulation method due to its limited range of applicability.

Multi-moment expansions are sometimes used for simulation in plasmas that are near equilibrium, such that the initial and perturbed particle distributions can be adequately described by the first few moments of the distribution function, as described in Section 2.1.5. However, while each successively higher moment captures more of the non-equilibrium statistics of a given distribution, the number of terms (and thus the computational cost) grows exponentially.

The particle-in-cell (PIC) method [Birdsall and Langdon, 1985] is a relatively well-established technique to model difficult problems in plasma physics. The technique has been applied to modeling a wide variety of problems: plasma expansion [Sarri et al., 2011], the Weibel instability [Stockem et al., 2010], laser-plasma interactions [Klimo et al., 2010], solar wind interactions [Omidì et al., 2002], magnetic reconnection [Daughton et al., 2006], and nonlinear Landau and cyclotron damping, among others.

The technique is extremely straightforward in principle. Instead of attempting to model a distribution as a phase-space fluid, the PIC method returns to the particle description of the plasma, approximating the population of particles with a population of *superparticles*. Each of these superparticles acts as if it were a cohesive cluster of particles all moving together under the influence of the same electric and magnetic fields. Since each particle artificially represents a very large number of physical particles, statistical fluctuations are typically much higher than they are in nature, the consequence being that noise can easily dominate the solution.

A PIC method consists of two basic parts, a particle mover and a field solver. Typical PIC implementations use FDTD to solve for the electric and magnetic fields, which are then interpolated at each particle location to move the particles. There is no fundamental reason the field solver must be FDTD, however, and there are some compelling reasons to consider other methods. The DG method discussed in the previous chapter, for instance, is high-order, local, and has all fields physically co-located.

As a high-order method, the accuracy of DG as a field solver is essentially a non-issue. Instead, errors in a DG-PIC scheme are nearly entirely dominated by the particle component. The strict locality of the scheme means that reconstructing a high-order interpolated solution at an arbitrary spatial point (which is necessary to

move the particles) requires only element-local information, eliminating inter-CPU communication for that stage of the PIC scheme. The co-location of all electric and magnetic fields is an advantage compared to field solvers utilizing staggered grids such as FDTD, particularly when simulating anisotropic media like magnetized plasmas. In order to maintain second-order accuracy in magnetized plasmas using FDTD, it is necessary to resort to extensive spatial averaging to ensure that the fields and currents are located at the same physical points in space. This averaging leads to artificial diffusion as well as other nonphysical effects. Further, the DG solution, while only solved for at interpolation points, is nonetheless defined everywhere within a given element.

It is natural to ask, then, whether using a high-order, high-quality field solver such as DG can lead to a general improvement of the PIC method, and according to early work on the DG-PIC method by *Jacobs and Hesthaven* [2006], it does, with particularly low levels of noise and excellent long-term stability properties compared to FDTD-PIC.

In this chapter, we discuss our extensions to the basic DG-PIC approach. We first review the basic scheme, followed by a description of our method for efficiently parallelizing both the projection and interpolation steps on generic unstructured meshes on a distributed (that is, not shared memory) system. We then describe our hybrid DG-PIC approach, where we treat a plasma as a combination of a cold background plasma modeled using a fluid approach and a hot electron component modeled using PIC. This approach is particularly useful for solving nonlinear wave growth problems in the Earth's magnetosphere, where a highly non-equilibrium, energetic electron population is frequently superimposed on a cold and (comparatively) dense background plasma.

Portions of this chapter have been published in *Physics of Plasmas* as *Foust et al.* [2011a].

4.1 PIC scheme

The basic particle-in-cell scheme samples a given particle distribution f using a collection of discrete particles at a set of points \mathbf{r}_i moving at velocity \mathbf{v}_i . Each particle

has an associated rest mass m_i and charge q_i . The particles are moved under the Lorentz force, where the required fields \mathbf{E} and \mathbf{B} , defined on a fixed mesh, must be interpolated to the particle locations:

$$\frac{d\mathbf{p}_i}{dt} = q_i \left[\mathbf{E}(\mathbf{r}_i) + \frac{1}{m_i \gamma_i} \mathbf{p}_i \times \mathbf{B}(\mathbf{r}_i) \right] \quad (4.1)$$

$$\frac{d\mathbf{r}_i}{dt} = \frac{\mathbf{p}_i}{m_i \gamma_i} \quad (4.2)$$

The particle velocity is related to the momentum, \mathbf{p} , by:

$$\mathbf{v}_i = \frac{\mathbf{p}_i}{\gamma m_i}, \quad (4.3)$$

and the Lorentz factor in terms of the momentum is:

$$\gamma = \sqrt{1 + \left(\frac{|\mathbf{p}_i|}{m_i c} \right)^2}$$

We use the momentum here to simplify the equations. Writing these equations in terms of the particle velocity would not (directly) yield an explicit ODE when relativistic effects are included.

The ensemble of particles interacts with the fixed mesh through a shape function S , chosen such that the ensemble of particles smoothly approximates the underlying distribution from a finite set of points.

$$\mathbf{J}(\mathbf{r}) = \sum q_i \mathbf{v}_i w_i S(|\mathbf{r} - \mathbf{r}_i|) \quad (4.4)$$

$$\rho(\mathbf{r}) = \sum q_i w_i S(|\mathbf{r} - \mathbf{r}_i|) \quad (4.5)$$

We have introduced the weight w_i , which is a measure of how many physical particles N at position \mathbf{r}_i moving with velocity \mathbf{v}_i are represented by the given numerical particle at index i . The window function $S(r)$ is a smoothing or weighting window

with some finite radius R . We use a standard window function:

$$S(r) = \begin{cases} \gamma \left(1 - \frac{r^\beta}{R^\beta}\right)^\alpha & 0 \leq r \leq R \\ 0 & r > R \end{cases}$$

The normalization constant γ is chosen such that volume integral of S integrates to exactly one. If the parameter β is equal to the number of spatial dimensions, the normalization constant has a simple closed-form solution. Otherwise, the normalization can be pre-computed using numerical quadrature. We have also found the triangle function ($\alpha = \beta = 1$) to perform adequately:

$$S(r) = \begin{cases} \gamma \left(1 - \frac{r}{R}\right) & 0 \leq r \leq R \\ 0 & r > R \end{cases}$$

Using an unstructured mesh complicates the PIC scheme, particularly when implemented in parallel. In the following sections, we discuss our approach to handling both the projection and interpolation steps in a parallel environment.

4.1.1 Projection step

The particle projection step involves computing an approximate current and charge density by accumulating weights onto a finite number of mesh points within a radius R about the particle position \mathbf{r}_i , weighted by the window S . Explicitly, for each particle i , we accumulate the currents and charges onto the grid as:

$$\mathbf{J}(\mathbf{r}_j) \leftarrow \mathbf{J}(\mathbf{r}_j) + q_i \mathbf{v}_i w_i S(|\mathbf{r}_j - \mathbf{r}_i|), \quad \{j \mid |\mathbf{r}_j - \mathbf{r}_i| \leq R\} \quad (4.6a)$$

$$\rho(\mathbf{r}_j) \leftarrow \rho(\mathbf{r}_j) + q_i w_i S(|\mathbf{r}_j - \mathbf{r}_i|), \quad \{j \mid |\mathbf{r}_j - \mathbf{r}_i| \leq R\} \quad (4.6b)$$

The task is made difficult by two complications. First, the set of mesh points must be queried quickly for all mesh points within a radius R of some given particle \mathbf{r}_i . Second, while a particle is represented by a single infinitesimal point and is therefore owned by only one CPU, the particle's shape function S may project across CPU

boundaries, so we must store not only the set of all mesh points owned by the current CPU but also all other mesh points on triangles owned by neighboring CPUs that a particle might be projected onto. We address both of these problems in order.

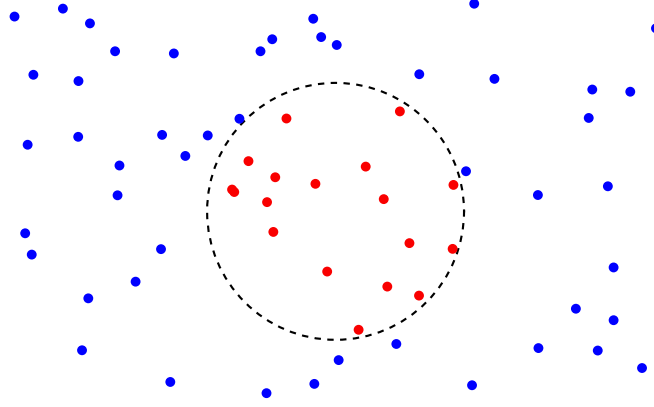


Figure 4.1: Illustration of the scattered point search problem. The dashed circle denotes the point search region. Mesh points within the search region are shaded red, while those external are in blue.

The mesh point search problem is illustrated in Figure 4.1. A naive algorithm requires order NM time, where N is the number of mesh points and M is the number of particles. We use two approaches to accelerate this procedure.

The first acceleration method uses a rectilinear grid into which points are binned. The tradeoff is identical to that encountered in hash tables. If the hash table is too coarse, then most of the time is spent performing a linear search through each bin. If the hash table is too fine, then space and time are wasted on empty bins.

The second acceleration technique uses a kd-tree. A kd-tree is a variant of a binary tree that is constructed by recursively subdividing a space into halfspaces, dimension by dimension. Each node of the tree corresponds to a cutting plane, whose cutting dimension is given by the depth of the given node. For example, if the root of the tree represents a division along the \hat{x} dimension, the second level would divide along the \hat{y} dimension, the third along the \hat{z} dimension, the fourth along the \hat{x} dimension, and so on. The left and right branches from a given node contain all points to the left or right of the cutting plane, respectively. Pseudo-code to build a kd-tree to sort mesh points is described in Algorithm 1.

Algorithm 1 Pseudo-code to build a kd-tree.

```

1: function BUILDKDTREE (points, depth)
2:   dim  $\leftarrow$  depth mod numDimensions
3:   points  $\leftarrow$  sortByDimension (points, dim)
4:   pivot  $\leftarrow$  round (length (points)/2)
5:   tree.point  $\leftarrow$  points(pivot)
6:   tree.dim  $\leftarrow$  dim
7:   tree.left  $\leftarrow$  BuildKdtree (points(1...pivot - 1), depth + 1)
8:   tree.right  $\leftarrow$  BuildKdtree (points(pivot + 1...end), depth + 1)
9:   return tree
10: end function

```

Once built, queries on the kd-tree only take logarithmic time. Pseudo-code for a log-time search for all points within a given radius is described in Algorithm 2. We note that while the kd-tree has very good formal complexity (log time), the bin search is in practice more efficient for uniform grids, since it has better locality of reference.

Algorithm 2 Pseudo-code to search a kd-tree for all points within a given radius.

```

1: function SEARCHKDTREE (tree, point, radius)
2:   if (tree.point(tree.dim) < point(tree.dim) + radius) then
3:     SearchKdtree (tree.right, point, radius)
4:   end if
5:   if (tree.point(tree.dim) > point(tree.dim) - radius) then
6:     SearchKdtree (tree.left, point, radius)
7:   end if
8:   if ( $\| \text{tree.point} - \text{point} \| < \text{radius}$ ) then
9:     OutputPoint (tree.point)
10:  end if
11: end function

```

The second problem to be solved is that particles violate the strict locality of the DG scheme, since a single particle's radius of influence may extend beyond not only a single element but across CPUs. The problem is illustrated in Figure 4.2. Our solution is to pre-build a list of indices of points on neighboring CPUs, all within a superset volume containing all points within a radius R of the local CPU's elements. This is computationally costly, since each CPU must broadcast its local set of points

to all other CPUs in turn, but this needs to be done only once as a pre-processing step.

Once these local indices and data structures are found, it is a simple matter to query for mesh points within a radius R . The actual accumulation step is performed using the Petsc library's generic vector scatter and accumulate, which computes the function $u_j \leftarrow u_j + v_i$, where \mathbf{u} represents one of the field unknowns (either \mathbf{J} or ρ), \mathbf{j} is a set of indices (which may be local or off-CPU), \mathbf{i} is a set of local indices, and \mathbf{v} is the window-weighted particle contribution.

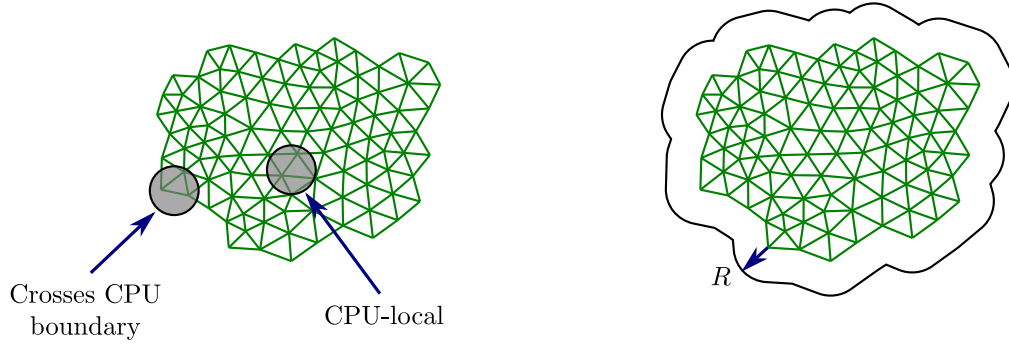


Figure 4.2: Illustration of the particle projection problem. **Left:** two particles' radii of influence are shown in gray. All mesh elements owned by the current CPU are drawn in green. The rightmost particle's range of influence is entirely contained within the local CPU. The leftmost particle crosses a CPU boundary. **Right:** CPU-local mesh elements showing its radius R bounding volume. Ghost indices of points on all neighboring CPUs contained within this superset are stored in a CPU-local table.

4.1.2 Interpolation step

In order to move the particles, we must reconstruct the solution at some arbitrary particle location \mathbf{r}_i using the DG basis:

$$\mathbf{E}(\mathbf{r}_i) = \sum_{j=1}^N \mathbf{E}_j \phi_j(\mathbf{r}_i) \quad \mathbf{H}(\mathbf{r}_i) = \sum_{j=1}^N \mathbf{H}_j \phi_j(\mathbf{r}_i)$$

While no closed-form expression exists for the interpolating basis (Lagrange polynomials) in two and three dimensions, we can nonetheless pre-compute the polynomial coefficients for a nodal set on a reference triangle, speeding the evaluation of the required fields at arbitrary particle locations. However, since the particles are free

to migrate between elements and even across CPUs, the question of which element owns a given particle must also be addressed. Since we are interested in beamlike and highly relativistic electron distributions characteristic of the magnetosphere, particles are not only free to move between CPUs but are expected to frequently do so.

Determining which element owns a given particle is a relatively simple proposition in exact arithmetic. For instance, in two dimensions, a triangular element's interior is the intersection of three halfspaces defined by the vertices $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}$ and surface normals $\{\hat{\mathbf{n}}_1, \hat{\mathbf{n}}_2, \hat{\mathbf{n}}_3\}$. A point is strictly within the triangle if $\hat{\mathbf{n}}_i \cdot (\mathbf{r} - \mathbf{r}_i) < 0 \ \forall i$. In finite-precision arithmetic, however, an ambiguity exists because of numerical round-off. In the worst cases, the procedure may conclude that no triangle owns a given point, leading to the disappearance of the particle from the scheme. Likewise, the procedure may conclude that more than one triangle owns a given point, leading to a doubling of that particle. Either case is extremely undesirable in a PIC simulation, since a single particle may represent a very large amount of physical charge.

On a shared-memory architecture, the ambiguity is easily resolved. Each processor can simply compare their inner products $\hat{\mathbf{n}}_i \cdot (\mathbf{r} - \mathbf{r}_i)$ to determine which triangle is closest to owning the given particle. In a distributed MPI architecture, however, this is an extremely costly proposition as it would require broadcasting information about every ambiguous particle at every timestep. We resolve this situation by identifying when a particle is within a circle of confusion where an ambiguity might exist, and then virtually pushing the particle away from the circle of confusion. This procedure is illustrated in Figure 4.3. Provided each CPU performs this procedure in an identical fashion, every CPU comes to the same conclusion about which element owns a given particle, effectively resolving the ambiguity without communication. This technique is similar to *simulation of simplicity*, described in detail in [Edelsbrunner and Mücke \[1990\]](#).

When a particle crosses an element boundary, a coarse search is necessary to first quickly determine which set of elements might become the point's new owner. Similar to the mesh search above, it is straightforward to generalize both the rectilinear grid accelerator and the kd-tree accelerator to hold element bounding boxes instead of infinitesimal points. For illustration purposes, we have plotted a sample grid with its

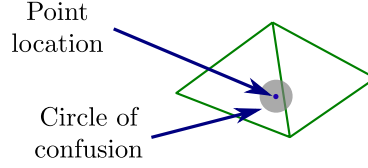


Figure 4.3: Illustration of the ambiguity introduced by finite-precision arithmetic. Near a boundary, an effective circle of confusion (shaded in gray) exists where the point's owning triangle cannot be unambiguously determined. Pushing the particle away from the boundary resolves the ambiguity.

kd-tree cutting planes overlaid in blue in Figure 4.4. We stop constructing the kd-tree at either a predefined maximum depth or when the number of elements contained within the kd-tree bounding volume is less than some predefined small number, say ten elements. Once the coarse search is performed, each individual element is queried in turn to find the true particle owner.

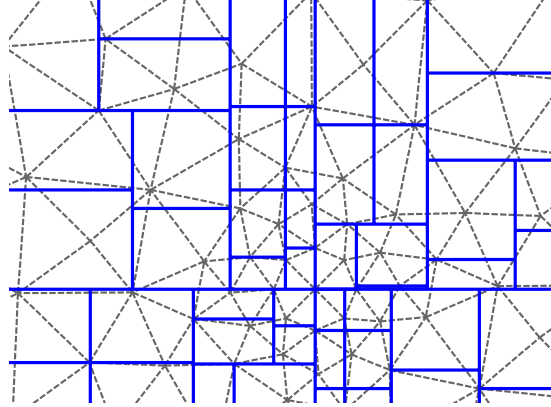


Figure 4.4: Illustration of a bounding box kd-tree. The blue lines indicate the locations of the cutting planes. The regions bounded by these cutting planes are leaf nodes of the kd-tree, each of which contains a list of pointers to any element whose bounding box overlaps this region.

4.1.3 Divergence correction

Time-domain solution techniques for Maxwell's equations typically ignore the two divergence constraints $\nabla \cdot \mathbf{D} = \rho$ and $\nabla \cdot \mathbf{B} = 0$. In a PIC simulation, however, ignoring the electric flux divergence can lead to significant errors. Consistency can be enforced by a number of techniques. We may opt to correct the electric field \mathbf{E} after every timestep by solving a Poisson problem. The constraints could also be explicitly

included by introducing a Lagrange multiplier. Either approach requires that a system of equations be solved in parallel every timestep, which unfortunately becomes extremely costly for large domains.

To avoid this problem, we use a divergence correction approach called *hyperbolic cleaning* [Munz et al., 2000; Jacobs and Hesthaven, 2006]. The hyperbolic cleaning method adds an additional scalar field ϕ to Maxwell's equations as:

$$\frac{\partial \mathbf{E}}{\partial t} - \frac{1}{\epsilon} \nabla \times \mathbf{H} + \chi \nabla \phi = -\frac{1}{\epsilon} \mathbf{J} \quad (4.7a)$$

$$\frac{\partial \mathbf{H}}{\partial t} + \frac{1}{\mu} \nabla \times \mathbf{E} = 0 \quad (4.7b)$$

$$\frac{\partial \phi}{\partial t} + \chi \left(\nabla \cdot \mathbf{E} - \frac{\rho}{\epsilon} \right) = -\sigma_\phi \phi \quad (4.7c)$$

It is clear that if $\nabla \cdot \mathbf{E} = \rho/\epsilon$, that is, if the divergence error is zero, then the scalar field ϕ is also zero. Similarly, in the limit as $\chi \rightarrow \infty$, the condition $\nabla \cdot \mathbf{E} = \rho/\epsilon$ is met exactly. The method effectively adds another propagating wave mode to the system, whose purpose is to propagate the divergence error out of the system at a velocity χ . The technique is quite similar to introducing a Lagrange multiplier to a mechanical system for the purpose of enforcing some physical constraint (e.g., a hinge, a stop, a wall, etc.). In this case, however, the extra equation introduced includes a time-varying operator $\partial/\partial t$ in addition to the standard terms. While this method only approximately enforces the constraint, in practice, if χ is a factor of five to ten larger than the free-space speed of light, the fundamental physics of the plasma are still captured [Munz et al., 2000].

Upwind flux for the augmented Maxwell system

Since we have added an additional mode and scalar field to Maxwell's equations, the numerical upwind flux also changes. The derivation of the upwind flux is given in Section B.3. Note that we have explicitly accounted for the possibility of a spatially-varying ϵ and μ to handle mixed simulations with both particles and mixed isotropic materials. Additionally, our expression corrects a minor sign error and a missing

factor of 1/2 in *Jacobs and Hesthaven* [2006]:

$$\hat{\mathbf{n}} \cdot (\mathbf{f}_E - \mathbf{f}_E^*) = \frac{1}{(Z^+ + Z^-)} (\alpha \hat{\mathbf{n}} (\hat{\mathbf{n}} \cdot [\mathbf{E}]) - \alpha [\mathbf{E}] - Z^+ \hat{\mathbf{n}} \times [\mathbf{H}]) + \quad (4.8a)$$

$$\frac{1}{2} \chi \hat{\mathbf{n}} ([\phi] - \hat{\mathbf{n}} \cdot [\mathbf{E}]) \quad (4.8b)$$

$$\hat{\mathbf{n}} \cdot (\mathbf{f}_H - \mathbf{f}_H^*) = \frac{1}{(Y^+ + Y^-)} (\alpha \hat{\mathbf{n}} (\hat{\mathbf{n}} \cdot [\mathbf{H}]) - \alpha [\mathbf{H}] + Y^+ \hat{\mathbf{n}} \times [\mathbf{E}]) \quad (4.8c)$$

$$\hat{\mathbf{n}} \cdot (\mathbf{f}_\phi - \mathbf{f}_\phi^*) = \frac{1}{2} \chi (\hat{\mathbf{n}} \cdot [\mathbf{E}] - [\phi]) \quad (4.8d)$$

As before, the bracketed notation $[\mathbf{u}]$ means the field difference at the face of a given element, $[\mathbf{u}] = \mathbf{u}^- - \mathbf{u}^+$, where \mathbf{u}^- denotes the field values at a face interior to a given element, and \mathbf{u}^+ denotes the field values at the face of the element's immediate neighbor. Z is the impedance, $(\mu/\epsilon)^{1/2}$. Y is the admittance, $(\epsilon/\mu)^{1/2}$. The term α is variable, where $\alpha = 1.0$ is full upwinding and $\alpha = 0.0$ is the central flux.

Boundary conditions

We impose a boundary condition on the scalar field ϕ such that it is consistent with PEC (perfect electrical conductor) boundary conditions imposed on the electric and magnetic fields. In physical terms, if a single charge were placed near a perfectly conducting boundary, the divergence correction field should establish an electric field \mathbf{E} consistent with placing an image charge of opposite sign mirrored across the boundary. The boundary condition is imposed during the calculation of the flux terms \mathbf{f}_E^* , \mathbf{f}_H^* , and \mathbf{f}_ϕ^* by setting the unknown fields \mathbf{E}^+ , \mathbf{H}^+ , and ϕ^+ “beyond” the boundary to:

$$\mathbf{E}^+ = -\mathbf{E}^-$$

$$\mathbf{H}^+ = \mathbf{H}^-$$

$$\phi^+ = -\phi^- - 2 \hat{\mathbf{n}} \cdot \mathbf{E}^-$$

We note that it is unnecessary to explicitly include a mirror charge opposite any charge near a conductive boundary as in *Jacobs and Hesthaven* [2006], provided the boundary conditions on ϕ are appropriately handled. Using our boundary conditions, the mirror charge emerges as a natural consequence of reflection of ϕ from the boundary.

4.1.4 Time integration

The maximum timestep of the augmented Maxwell system is limited by the divergence correction field velocity χ . Since χ should be set to on the order of ten times the speed of light in free space for acceptable accuracy, the augmented Maxwell system is also stiff, but in the DG (spatial) operator rather than in the plasma currents. We use an approach identical to that described in *Jacobs and Hesthaven [2009]*. As in Section 3.4.3, we split the right-hand side into a sum of two components:

$$\frac{d\mathbf{u}}{dt} = f^{[\text{im}]}(\mathbf{u}) + f^{[\text{ex}]}(\mathbf{u}) \quad (4.9)$$

In this case, however, $f^{[\text{im}]}(\mathbf{u})$ is the DG operator itself:

$$f^{[\text{im}]} \left(\begin{bmatrix} \mathbf{E} \\ \mathbf{H} \end{bmatrix} \right) = \begin{bmatrix} 0 & \frac{1}{\epsilon} \nabla \times \\ -\frac{1}{\mu} \nabla \times & 0 \end{bmatrix} \begin{bmatrix} \mathbf{E} \\ \mathbf{H} \end{bmatrix}$$

Since the upwind and central fluxes are both linear, the DG operator is also linear and can be written simply as a large, sparse matrix, \mathbf{L} , times the stacked vector of all unknowns \mathbf{u} :

$$f^{[\text{im}]}(\mathbf{u}) = \mathbf{L}\mathbf{u}$$

The matrix \mathbf{L} is large but very sparse and diagonally dominant. For context, we show a typical sparsity pattern for the DG operator matrix in Figure 4.5. The off-diagonal entries are the weights for the flux differences at the faces between elements while the on-diagonal entries are the mass and stiffness matrices associated with each element.

At each Runge-Kutta stage, a system of the following form must be solved:

$$\left[\mathbf{I} - (\Delta t) a_{ii}^{[\text{im}]} \mathbf{L} \right] \mathbf{u}^{(i)} = \mathbf{r},$$

where $a_{ii}^{[\text{im}]}$ is the weight for the i^{th} implicit Runge-Kutta stage (See Appendix C and Section 3.4.3) and \mathbf{r} collects the right hand side terms for that stage. When

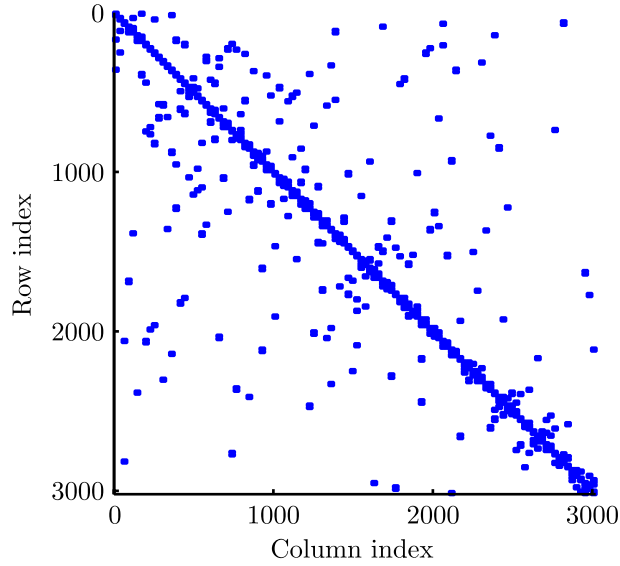


Figure 4.5: Sparsity pattern for the DG operator matrix L for a 2D square domain discretized into 112 triangles using a first-order (linear) DG discretization. Diagonal entries represent the intra-element operations (the mass and stiffness matrices), while off-diagonal entries represent the inter-element operations (the flux contribution at the element faces).

the matrix is small, we solve this system with the Mumps LU solver [Amestoy *et al.*, 2001, 2006]. For larger systems it is necessary to use iterative methods. We find that the BiCGSTAB (BiConjugate Gradient STABilized) iterative method [van der Vorst, 1992] with the Euclid preconditioner [Hysom and Pothen, 2000, 1999] performs exceptionally well for this system, acceptably converging in fewer than 10 iterations for moderately sized problems.

The particles are moved in the explicit part of the IMEX step, that is, no global solve is required for the particle mover. We note that integrating the particle step implicitly would be prohibitively expensive since the operator matrix would change with every iteration.

4.1.5 Hybrid approach

Representing the entire distribution of both ions and electrons with particles can be extremely slow. For certain types of problems, it is possible to reduce this cost by using a so-called *hybrid* particle-in-cell method. If the number of hot particles N_{hot} is much smaller than the number of cold particles N_{cold} , this approximation

can significantly reduce the computational requirements (and consequently, the noise levels) in PIC simulations.

In this approach, we split the distribution function into two parts, a hot component handled using PIC and a cold component using a fluid or cold plasma method.

$$f = f_{\text{cold}} + f_{\text{hot}}$$

By linearity, it follows that the currents and charges from these two particle populations also sum:

$$\begin{aligned}\rho &= \rho_{\text{cold}} + \rho_{\text{hot}} \\ \mathbf{J} &= \mathbf{J}_{\text{cold}} + \mathbf{J}_{\text{hot}}\end{aligned}$$

An additional assumption we make is that there is no continuity law between the two distributions, that is, a cold particle cannot become a member of the hot population and vice-versa. Typically, this is a safe approximation to make if either the hot particle number density N_{hot} is much less than the cold background density N_{cold} , or if the two populations of particles are sufficiently separated from each other in phase space. Under these assumptions, we can directly evaluate the cold plasma current via the method described in Section 3.2.2 and evolve a population of hot particles using the PIC technique.

One minor complication is that while \mathbf{J}_{cold} is evaluated explicitly by our cold plasma technique (Section 3.2.2), the cold space charge ρ_{cold} is not. For waves propagating obliquely to the magnetic field \mathbf{B}_0 , space charge can exist even in the cold plasma approximation. Such a space charge is supported by a time-varying wave electric field rather than a pressure (which is zero in the cold plasma approximation). If we were to make the naive assumption that ρ_{cold} is identically zero everywhere, this assumption would lead to an inconsistency for obliquely-propagating waves when evolving the divergence correction field ϕ via Equation 4.7c.

Thus, for obliquely-propagating waves, it is also necessary to evolve an auxiliary

PDE in order to determine the cold plasma space charge ρ_{cold} :

$$\frac{\partial \rho_{\text{cold}}}{\partial t} + \nabla \cdot \mathbf{J}_{\text{cold}} = -\alpha_{\rho} \rho$$

The loss term α_{ρ} can be set to some small positive value to help ensure that large deviations from quasi-neutrality do not accumulate with time.

4.2 Validation

In this section, we establish the accuracy of the DG-PIC technique and the DG-PIC hybrid technique by comparison to analytical solutions. While plasma physics suffers from a paucity of explicit solutions to many problems of interest, there nonetheless exist some, valid for small signal amplitudes or under other approximations (e.g., the cold plasma approximation), that can be used to validate the basic technique.

We are primarily concerned with the behavior of waves in a plasma, so we restrict our discussion to propagation, growth, and damping of electrostatic and electromagnetic waves for which analytical solutions exist. The properties of electromagnetic wave propagation in the DG-PIC scheme are at present relatively poorly documented. We seek to address this problem in this section. We begin with a short investigation of the basic electrostatic behavior of the scheme in order to establish its long-term stability properties.

4.2.1 Grid heating

PIC simulations often exhibit a slow rise in the total energy in the system caused by a weak instability called *grid heating*. Grid heating represents a failure of the underlying fixed grid to represent high spatial frequencies in the particle population. Essentially, the error is caused by aliasing due to failure to sufficiently sample fluctuations on the scale of the Debye length, manifesting as an exponential, nonphysical energy rise in the system as a function of time.

In order to illustrate this failure, we choose parameters that are likely to give rise to very strong grid heating. We choose an electron temperature less than an order of magnitude smaller than in the next section and discretize the same domain using

only 25 elements instead of 100. Under these conditions, we have approximately 23 Debye lengths per element, that is, the simulation is significantly under-resolved and we would expect the artificial energy rise to be very large. Additionally, we initialize the system with relatively few particles, in the range of 32 to 128 per element, in order to ensure that grid heating manifests over a reasonable time scale. The results are shown in Figure 4.6, where we plot the total energy (the sum of the particle kinetic energy and the potential energy in the electric field) as a function of time for a number of different parameters. The variable n denotes the number of particles per element. R denotes the particle radius as a factor of the element size, that is, $R = 1$ means the particle cloud radius is the same size as one element. N_p denotes the polynomial order of accuracy. Increasing the number of particles decreases grid heating slightly due to the higher smoothness of the original distribution. Increasing the particle radius has the same effect for the same reason. The most dramatic reductions in the magnitude of the artificial energy rise, however, are realized by increasing the polynomial order (and consequently, the number of unknowns per unit length) from $N_p = 3$ to $N_p = 4$. Similar results are obtained by increasing the number of elements from 25 to 50. If we increase all three parameters, the scheme exhibits practically zero energy rise over 100 plasma oscillations.

4.2.2 Conservation of momentum

In order to investigate global conservation of momentum, we initialize the system with a cold, drifting electron plasma in a neutralizing (constant opposing current, constant positive charge) background. A cold, drifting plasma is expected to be physically stable for all wavenumbers. Failure of either conservation of energy or conservation of momentum would manifest as an instability in this simulation, driven by inaccuracies as the particles drift over the finite grid. We choose the drift velocity such that a given particle makes approximately 16 transits of the domain over a period of 200 plasma oscillation periods. Again, we use a relatively low number of particles on a coarse grid, that is, the simulation is significantly under-resolved.

The results are shown in Figure 4.7. Grid heating is present, manifesting as a very rapid initial phase of exponential growth in the total energy of the system. For

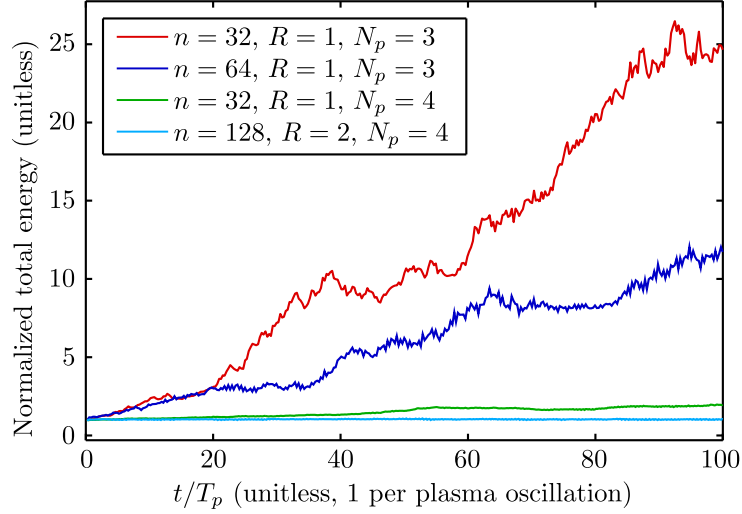


Figure 4.6: Plot of the normalized total energy (TE = KE + PE) for a randomly-sampled Maxwellian electron plasma against an immobile, neutralizing background. Four cases are shown for various selections of the number of particles per element n , the relative particle radius R (as a fraction of the average element size), and the polynomial order N_p . These simulations are deliberately under-resolved in order to clearly show the effects of grid heating, with on the order of 20 Debye lengths per finite element.

reference, we plot this total energy in the left panel. The first phase of growth (from approximately $t = 0$ to $t = 60$ plasma periods) is dominated by this exponential energy rise. Later times are dominated by individual particle fluctuations, a consequence of using too few particles. The instantaneous total momentum is dominated by small-scale fluctuations due to the small number of particles. Initially these fluctuations are on the order of the momentum of a single particle cloud. As the instability progresses, however, the beam is heated, and these fluctuations in the instantaneous total momentum grow larger, just as we would expect based on the corresponding rise in the total kinetic energy. A global failure of conservation of momentum manifests itself as a slow change in the average momentum of this beam as it transits through the domain due to viscous losses or artificial gains, that is, the beam, en masse, either slows down due to artificial numerical friction or speeds up for a similar reason.

In order to more clearly illustrate these long-term trends, we filter the instantaneous momentum using a rectangular window of width equal to 4 plasma oscillation periods. The results are shown in the right panel, plotted as the fractional change

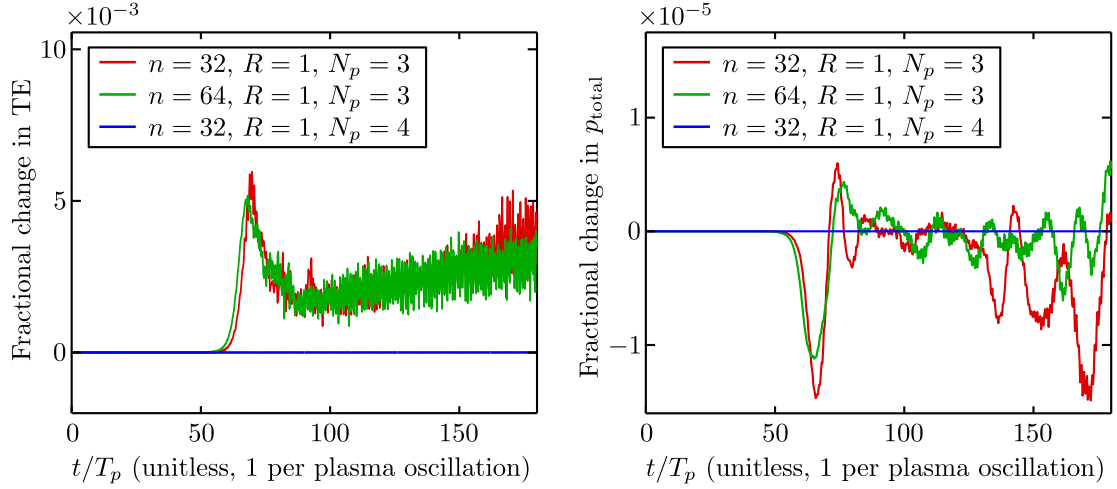


Figure 4.7: Left: Fractional change in the total energy ($TE = KE + PE$) for a DG-PIC simulation of a cold beam drifting against a neutralizing background. Two unstable and one effectively stable case are shown. **Right:** The fractional change in the time-averaged total momentum for the same simulation, plotted as a function of time. The time-averaged total momentum is calculated by windowing the instantaneous total momentum with a rectangular window of size equal to 4 plasma oscillation periods.

in the average total momentum as a function of time. Despite the fact that momentum is not explicitly conserved by the DG-PIC method, it nonetheless performs adequately, showing only small fluctuations in the average momentum and no significant long-term loss or gain trends, even in the very unstable cases. Further, increasing the polynomial order from $N_p = 3$ to $N_p = 4$ stabilizes the simulation. No appreciable energy rise or change in the average total momentum is observed.

4.2.3 Landau damping

One of the simplest waves in a plasma is the electrostatic oscillation at the plasma frequency. If the plasma has a nonzero temperature, the plasma oscillation can become a propagating mode. For a one-component plasma, the dispersion relation is given by the following approximate expression, where ω_r is the wave frequency:

$$\omega_r^2 = \omega_p^2 + 3k^2 v_{th}^2 \quad (4.10)$$

The thermal velocity v_{th} is:

$$v_{th} = \sqrt{\frac{k_B T}{m}}$$

Alternatively, the dispersion relation can be computed exactly by direct evaluation of the linearized Vlasov equation as discussed in Section 2.2.2 and summarized in Equations 2.27 and 2.28 for damping and growth, respectively.

For our validation, we load a periodic domain with a Maxwellian distribution with a finite electron temperature T and compare the oscillation frequency and damping rate to those expected from linearized Vlasov theory. Noting that the longitudinal electron mode cannot affect the distribution perpendicular to the direction of propagation, we use a purely one-component Maxwellian distribution function:

$$f(v) = N \left(\frac{m}{2\pi k_B T} \right)^{\frac{1}{2}} \exp \left(-\frac{mv^2}{2k_B T} \right) \quad (4.11)$$

The linear Landau damping rate is derived by approximately solving Equation 2.27 under the assumption of small loss [Landau, 1946]. Assuming that the frequency ω has a very small imaginary component ω_i representing loss, the singularity at $v = \omega/k$ is handled by integrating along a complex contour, where in typical derivations, the contour is chosen to take a half-circle detour under the singularity, as opposed to our full circle contour shown in Figure 2.3. Expanding the denominator in a series and further assuming that $k\lambda_D \ll 1$ yields an approximate expression for the damping rate:

$$\omega_i = -\sqrt{\frac{\pi}{8}} \exp \left(-\frac{3}{2} \right) \frac{\omega_p}{(k\lambda_D)^3} \exp \left(-\frac{1}{2(k\lambda_D)^2} \right) \quad (4.12)$$

This expression only holds for very small damping rates relative to the plasma oscillation frequency. As we show below, this expression is highly inaccurate over the parameter ranges commonly chosen in PIC simulations, necessitating an exact evaluation of Equation 2.27 for accurate comparison with theory. For high damping rates, the correct contour is given by Figure 2.3, so the singularity then contributes a factor

of $2\pi j$ as opposed to πj in the small ω_i limit.

Direct evaluation of the electrostatic dispersion relation

We evaluate the integral in Equation 2.27 at some ω and k by numerically integrating over some appropriate range of velocities. The trapezoidal rule works well for this integration, but other approaches could also be used. The dispersion relation is then solved as an optimization problem. In the half-plane $\omega_i < 0$, the dispersion relation as a function of ω is smooth and well-behaved within the region of interest, so we can solve the system by minimizing:

$$g(\omega) = \log |\epsilon(\omega, k_0)| \quad (4.13)$$

using a standard direct search method for unconstrained optimization over (ω_r, ω_i) . We have found that while the function is smooth and exhibits a clear minimum for the Maxwellian distribution, the minimum can exist in a long, narrow valley in (ω_r, ω_i) space when the loss rates are very small. Using the logarithm reduces the relative narrowness of the valley, and this in turn helps reduce the overall number of iterations in the optimization routine. The true root should be in the vicinity of real and imaginary frequencies given by Equations 4.10 and 4.12 (the lossless longitudinal wave frequency ω_{r0} and damping rate ω_{i0} , respectively), so we choose those frequencies as our initial guess $(\omega_{r0}, \omega_{i0})$.

Validation

We discretize a periodic domain using 100 finite elements and a fourth order polynomial basis. To reduce the possibility of confounding numerical effects, we initialize the system with a very large number of particles, 1024 per element for low damping rates and 8192 for very high damping rates. Based on our grid heating tests, this should be more than adequate to control grid heating for the duration of the simulation. The particles are loaded using a quiet start (ordered v - x space [*Birdsall and Langdon, 1985*]) sampling. Each particle has the same charge-mass ratio but a different cloud weight, so particles on the long tail of the Maxwellian contribute proportionally less. The spacing in v space is chosen narrow enough to eliminate the

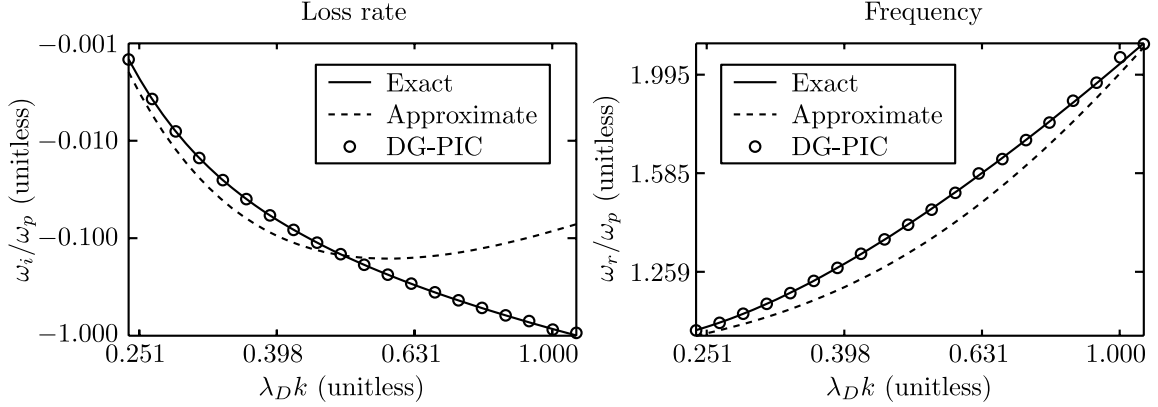


Figure 4.8: **Left:** Log-log plot showing the approximate (Landau) damping rate, exact damping rate, and the measured damping rates from DG-PIC simulation for a Maxwellian distribution. **Right:** Same but for oscillation frequency.

multi-beam instability over the simulation time of interest. We simulate the system over a range of $\lambda_D k$, from approximately 0.25 (weakly damped, $\omega_i/\omega_p \simeq -0.002$) to 1 (strongly damped, $\omega_i/\omega_p \simeq -1.0$).

The system is initialized with very small sinusoidal perturbation on the electron position, providing the initial (real) wavenumber k . The system is run long enough to observe significant damping. The frequency of oscillation is determined by peak detection, and the damping rate is determined by using an L1 norm log-linear fit to the peaks.

The results are summarized in Figure 4.8. We show the Landau prediction and the damping rate as predicted by exact numerical evaluation of the plasma dispersion relation. The measured PIC simulation results are superimposed as discrete circles. While PIC simulation and the exact solution agree quite well, the Landau linearization (Equation 4.12) only agrees well at extremely low damping rates.

4.2.4 Cold plasma

The cold plasma approximation is a useful point of reference to investigate the basic electromagnetic behavior of the DG-PIC scheme, since the approximation yields explicit, closed-form solutions for the dispersion relation. A particularly compact derivation is derived in Section 2.2.1. In a PIC simulation, a cold plasma is approximated by setting all initial particle velocities to zero. In other words, we sample the distribution

function:

$$f(v_x, v_y, v_z) = N\delta(v_x, v_y, v_z)$$

Charge neutrality is enforced by either using particles for both ions and electrons, or by imposing a background, immobile charge present over all space. As is described in Section 3.5, the numerical dispersion relation is computed by pinging the system with a spatially and temporally-compact pulse, recording the fields as a function of space and time, and then Fourier transforming into ω - k space.

Since all particles in this particular simulation have zero initial velocity, the only other PIC parameters we must choose are the number of particles per meter per dimension and the individual particle radius. The particle radius must be chosen large enough so that it encloses some nonzero number of static grid points. If the radius is too small, a particle apparently disappears and reappears as it traverses the untracked space between grid points, resulting in an instability. If the radius is too large, the particles instantaneously affect the electric and magnetic fields on faraway grid points, which could manifest either as an instability or as a departure from expected theoretical behavior.

Since a numerical instability corrupts any results to the point of uselessness, we restrict ourselves to only a brief discussion of the finite particle size on the propagating mode characteristics. Based on the results in the previous section, we first choose particles with radius approximately equal to the grid size. The electron number density and background magnetic field are identical to those used in Section 3.5: $N = 3.2 \times 10^{10} \text{ m}^{-3}$ and $B_0 = 4.34 \times 10^{-5} \text{ T}$. Figure 3.14 shows the theoretical cold plasma dispersion diagram for these parameters. Our results are referenced against this diagram. The results are shown in Figure 4.9. The propagating modes match quite closely with theory, effectively identical for $k < 0.03 \text{ m}^{-1}$ and diverging slightly for higher wavenumbers.

If we increase the particle radius to a much higher value, say five times the grid spacing, the scheme fails. We show these results in Figure 4.10. The propagating modes are severely distorted, showing both spectral shifts and modes with negative

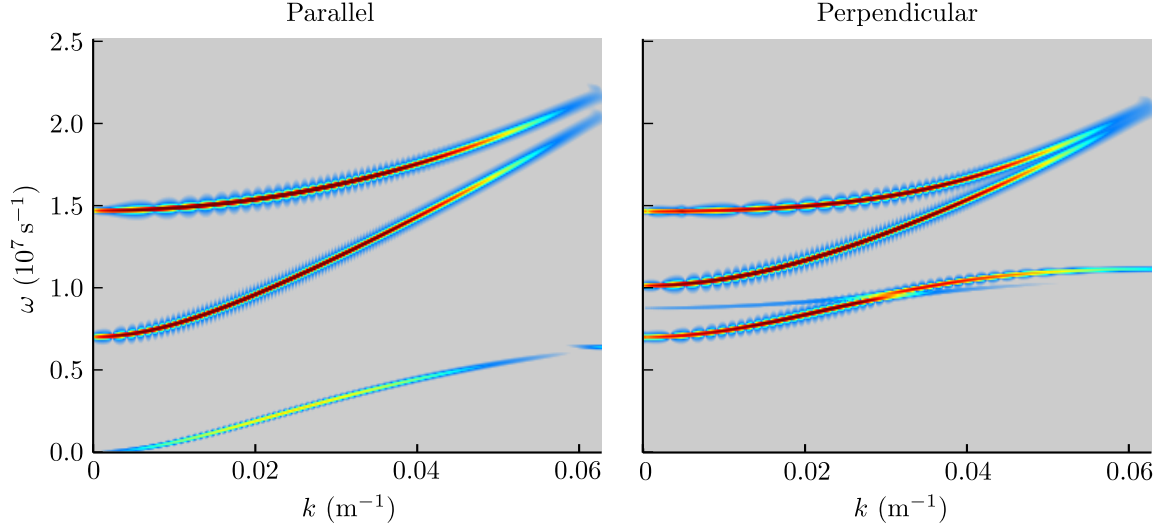


Figure 4.9: Numerical dispersion relation for a magnetized plasma using the DG-PIC scheme with small particles (radius equal to the element size), showing physical behavior. The left plot is for propagation parallel to the background magnetic field while the right is for propagation perpendicular to it. The parameters are identical to those used in Figure 3.14.

group velocities for some range of spatial frequencies, e.g., the uppermost branch in the spatial frequency range $0 < k < 0.035 \text{ m}^{-1}$.

4.2.5 Cyclotron damping and growth

Landau damping, as discussed in Sections 2.2.2 and 4.2.3, is only one means by which energy can be exchanged between a particle distribution and a wave. As discussed in Section 2.2.3, an electromagnetic wave can also exchange energy with a particle distribution through the cyclotron resonance interaction, leading to either damping or growth, depending on the polarization of the wave and the particular character of the distribution function.

For sufficiently unperturbed particle distributions, we can derive a dispersion relation (Equation 2.29), yielding the oscillation frequency and the spatial or temporal damping rate for a given electromagnetic wave. In principle, Equation 2.29 can be solved exactly, but such a solution is complicated in practice by the infinite sum over the resonances indexed by n . A particularly useful simplification is to assume the wave propagates parallel to the background magnetic field. In this case, each term in the sum over n vanishes except for the $n = \pm 1$ resonances, yielding Equation 2.30.

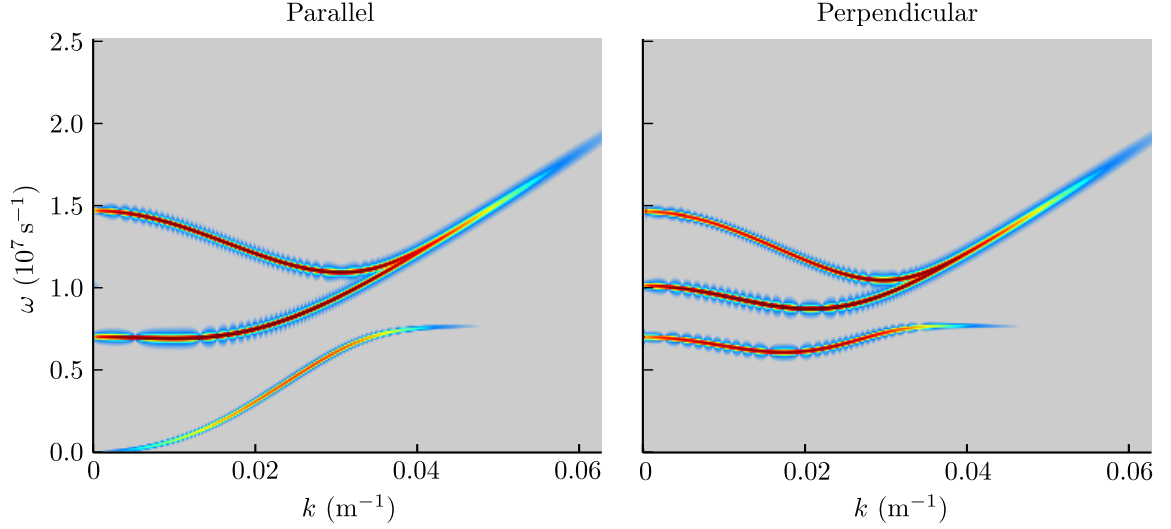


Figure 4.10: Numerical dispersion relation for a magnetized plasma using the DG-PIC scheme with large particles (radius five times the element size), showing nonphysical behavior. Each propagating branch is severely distorted, with some branches and ranges of frequencies exhibiting negative group velocities (when $\partial\omega/\partial k$ is negative). The left plot is for propagation parallel to the background magnetic field while the right is for propagation perpendicular to it. The parameters are identical to those used in Figure 3.14.

This simplification once again introduces a non-integrable singularity, which we handle in the same way as in Section 2.2.2, yielding the following dispersion relation, valid for parallel propagation in a hot magnetized plasma.

$$1 - \frac{c^2 k_{\parallel}^2}{\omega^2} - \frac{q^2}{m\epsilon_0\omega^2} \left[\int_{-\infty}^{\infty} \frac{g(v_{\parallel})}{v_{\parallel} - v_{\text{res}}} dv_{\parallel} + 2\pi j H(-\text{Im}\{\omega\}) g(v_{\text{res}}) \right] = 0, \quad (4.14)$$

with $g(v_{\parallel})$ defined in Equation 2.31, $v_{\text{res}} = (\omega \pm \omega_c)/k_{\parallel}$, and $H(x)$ is the unit step function. Note that the unit step function is used here to switch between the gain and loss cases, i.e., when the singularity crosses the real axis and begins to deform the integration contour as illustrated in Figure 2.3. In some cases, it is useful to evaluate this dispersion relation for a particle distribution split into a hot and cold part, i.e.:

$$f(v_{\parallel}, v_{\perp}) = f_h(v_{\parallel}, v_{\perp}) + \frac{N_c}{2\pi v_{\perp}} \delta(v_{\perp}) \delta(v_{\parallel})$$

Which leads to the expression:

$$1 - \frac{c^2 k_{\parallel}^2}{\omega^2} - \frac{q^2}{m\epsilon_0\omega^2} \left[\int_{-\infty}^{\infty} \frac{g_h(v_{\parallel})}{v_{\parallel} - v_{\text{res}}} dv_{\parallel} + N_c \frac{\omega}{\omega \pm \omega_c} + 2\pi j H(-\text{Im}\{\omega\}) g_h(v_{\text{res}}) \right] = 0$$

Either of these expressions can be evaluated numerically as described in Section 4.2.3. Similar to Landau damping, this expression can be further approximated using the assumption of small growth or damping ω_i relative to the wave frequency ω_r , yielding Equation 2.32.

Validation

We validate the hybrid DG-PIC scheme for cyclotron growth and damping by direct comparison to linearized Vlasov theory, both to the approximate linearized cyclotron growth/damping approximation and to the exact solution to the hot plasma dispersion relation as described previously.

We discretize a 1D periodic domain into 8 grid elements with a fourth-order DG approximation and fill this domain with a cold electron plasma with number density N_{cold} . Within each element, we also sample an unstable hot electron distribution with density N_{hot} in (v_x, v_y, v_z) space using 801 particles in the direction parallel to the magnetic field and 11 particles in each dimension perpendicular to it. The background magnetic field and the free spatial domain are in the same direction, that is, we simulate parallel propagation of whistler mode waves. The electron cyclotron frequency and cold electron plasma frequency are 13.6 kHz and 27.3 kHz, respectively. We use a bi-Maxwellian loss-cone distribution [Katoh and Omura, 2006] of the form:

$$f_{\text{hot}} = \alpha N_{\text{hot}} \exp\left(-\frac{v_{\parallel}^2}{2v_{\text{th},\parallel}^2}\right) \frac{1}{1-\beta} \left[\exp\left(-\frac{v_{\perp}^2}{2v_{\text{th},\perp}^2}\right) - \exp\left(-\frac{v_{\perp}^2}{2\beta v_{\text{th},\perp}^2}\right) \right] \quad (4.15)$$

The parameter α is chosen so the distribution integrates to N_{hot} . The remaining parameters are: $v_{\text{th},\parallel} = 0.15c$, $v_{\text{th},\perp} = 0.20c$, and $\beta = 0.5$.

We simulate this system for two choices of hot particle densities, one relatively small with $N_{\text{hot}} = 0.05N_{\text{cold}}$, and one large with $N_{\text{hot}} = 0.5N_{\text{cold}}$. As with Landau damping, the approximate linearized growth/damping rate given by Equation 2.32

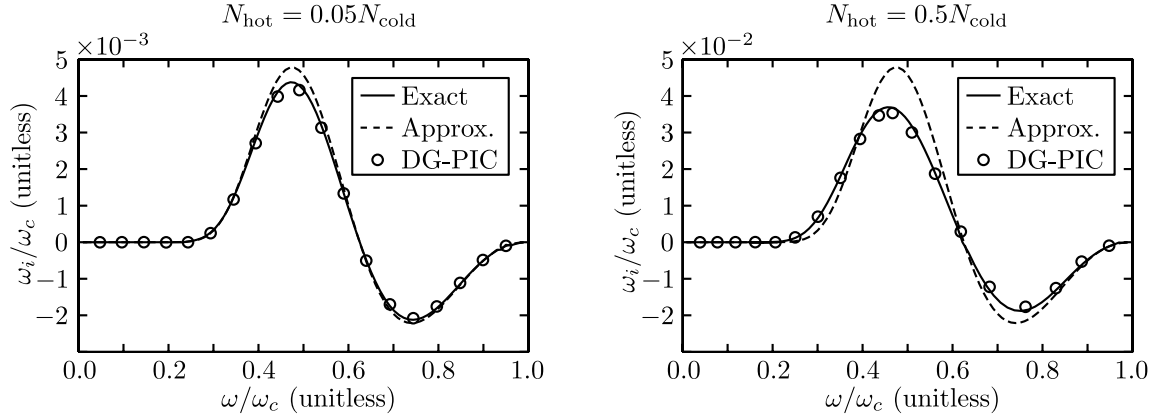


Figure 4.11: Normalized (with respect to the electron gyrofrequency ω_c) electromagnetic cyclotron wave growth/damping rates plotted as a function of normalized (again with respect to ω_c) of the wave frequency ω . The plots show the exact damping rates calculated numerically from Equation 4.14, the approximate damping rates expected from Equation 2.32, and the results from a hybrid DG-PIC simulation. The left plot is for a plasma with few hot particles ($N_{\text{hot}} = 0.05N_{\text{cold}}$) while the right is for many ($N_{\text{hot}} = 0.5N_{\text{cold}}$).

is expected to be a close match to reality only for very small values of ω_i , which we demonstrate in our results.

We initialize the domain with a very small whistler-mode wave at a fixed wavelength and allow the system to evolve long enough to measure the resulting temporal wave frequency and temporal growth rate. We summarize our results in Figure 4.11. For low damping and growth rates (left panel), the exact, approximate, and DG-PIC results all match quite closely. For higher (but still linear) damping and growth rates (right panel), the approximate expression begins to diverge. In both cases, the DG-PIC results and the exact dispersion relation agree closely, validating the basic hybrid DG-PIC scheme for cyclotron growth and damping.

4.3 Summary

In this chapter, we have described an efficient, parallel implementation of a hybrid PIC scheme on unstructured grids in the DG framework. We have made a number of contributions to the basic DG-PIC scheme, with particular emphasis on its efficient parallel implementation on distributed clusters and its performance for simulating electromagnetic wave propagation and growth.

Specifically, we have described a method to accelerate the element containment search by use of a pre-computed plaid grid or kd-tree, additionally resolving any geometric ambiguities that may arise without resorting to voting strategies that would require inter-node communication. We have also devised a method to accelerate the particle projection step on completely unstructured grids, again specifically targeting distributed environments where inter-node communication costs may be high.

Our hybrid DG-PIC approach incorporates both fluid and particle descriptions of a plasma in one scheme. The hybrid approach reduces the noise and computational requirements when simulating problems for where there exists a clear separation between a background, cold magnetized plasma (simulated using the fluid technique) and a hot unstable component (simulated using the PIC technique). We have demonstrated the correctness of this approach by direct comparison to theory, showing that the scheme correctly reproduces cyclotron growth rates predicted by the linearized Vlasov equation.

Chapter 5

Scattering from ionospheric disturbances

Having established the basic validity and flexibility of the DG technique for simulating wave propagation in cold magnetized plasmas, we now turn our attention to simulating physical problems of interest.

This chapter discusses the application of the DG technique to modeling VLF wave scattering from lightning-induced ionospheric disturbances. This chapter extends the work of *Lehtinen et al.* [2010], who used a linearization to estimate the VLF scattered fields from relatively mild lightning-induced ionospheric disturbances. Since our technique is not limited to mild perturbations, we can investigate the much higher limits (for very strong lightning discharges) of this particular scattering mechanism and its impact on VLF wave propagation in the Earth-ionosphere waveguide.

Portions of this chapter have been published in the Journal of Geophysical Research as *Foust et al.* [2011b].

5.1 Background

Lightning discharges can modify the ionosphere via a number of different mechanisms. Due to the impulsive, energetic nature of a lightning return stroke, a large electromagnetic pulse (EMP) is radiated from a lightning channel. The fields associated with this EMP are intense enough to cause significant heating and ionization of the ionosphere

above the discharge [Inan *et al.*, 1991; Cheng *et al.*, 2007b; Marshall *et al.*, 2008a, 2010]. A number of secondary effects are also present. For instance, the amount of charge removed by a lightning return stroke creates a large quasi-electrostatic (QE) field above a lightning discharge, which can further modify the levels of ionospheric ionization [Pasko *et al.*, 1995, 1998]. In either the QE or the EMP cases, the magnitude of the change in the ambient electron density can be on the order of a few percent to over a hundred percent, peaking in an altitude range of approximately 80 to 100 km and extending over radial distances of over 100 km [Inan *et al.*, 1991; Taranenko *et al.*, 1993; Cheng *et al.*, 2007b; Marshall *et al.*, 2010]. Due to the abrupt onset of such disturbances, their altitudes near the VLF reflection height, and their relatively wide spatial extent, these ionospheric disturbances are expected to modify the properties of VLF waves propagating within the Earth-ionosphere waveguide. This modification can be observed by a ground-based receiver as a phase or magnitude perturbation on a received narrowband signal observed shortly after a causative lightning discharge. Indeed, this connection was observed quite early [Armstrong, 1983]; however, the exact nature of the disturbances and their effects on received VLF signals was not elucidated until later modeling efforts [Inan *et al.*, 1991; Taranenko *et al.*, 1993; Cheng *et al.*, 2007b; Marshall and Inan, 2010].

We discuss one particular class of VLF signal disturbances, termed *early* VLF events. Early VLF events are abrupt changes in the amplitude of a received narrowband VLF signal following a lightning discharge. They are characterized by their abrupt onset (<20 ms) after a causative lightning discharge (hence the term *early*) and their relatively slow recovery time (10-100 seconds). Initially, a number of different causative mechanisms were postulated [Inan *et al.*, 1991; Pasko *et al.*, 1998; Moore *et al.*, 2003]. Recent modeling efforts have shown that EMP-induced ionospheric disturbances may be consistent with some such events [Cheng and Cummer, 2005; Cheng *et al.*, 2007b; Marshall and Inan, 2010]. A recent review article [Inan *et al.*, 2010] summarizes what is known (as of 2010) of these lightning-ionosphere interactions, among others.

Previous work by Marshall and Inan [2010] used a 2D finite difference model to show that the EMP from strong (but realistic) lightning discharges, with E_{100} (defined

in *Uman and McLain* [1970] as the magnitude of the electric field as observed on the ground at 100 km from a discharge) in the range of 7 to 40 V/m, can perturb the ambient electron density enough to cause measurable (>1 dB) perturbations on the amplitude of a received narrowband VLF signal. Later work by *Lehtinen et al.* [2010] used the Born approximation to approximate the scattered field in a medium of homogeneous, horizontally stratified layers. This work revealed the full three-dimensional structure of the scattered field from a lightning EMP-induced density perturbation for a range of incidence angles. The technique used in *Lehtinen et al.* [2010], however, assumes that the scattered field is much smaller than the incident field and as such is unsuitable for the whole range of density perturbations considered in *Marshall and Inan* [2010]. In addition, the technique cannot account for multiple scattering or modes propagating exactly parallel to the stratified layers, necessitating the use of a more general solution method for strong perturbations. The magnitude, angular extent, and shape of the scattered field and amount of transverse variation for intense, spatially complicated disturbances were previously unknown.

We extend the work of *Lehtinen et al.* [2010] and *Marshall and Inan* [2010], using a fully three dimensional continuum electromagnetic DG solver to solve for the VLF scattered fields from an EMP-induced ionospheric disturbance. As a continuum, time-domain method, backscattering and multiple scattering are implicitly handled. The only significant limitation is that the wavelengths of interest must be sufficiently sampled. For the case considered in this chapter, the free-space wavelength is equal to 12.5 km but in the ionosphere drops to less than 2 km for propagation parallel to the magnetic field.

5.2 Scattering model

Waves radiated by an antenna at VLF frequencies (3-30 kHz) are effectively guided between the ground and the conductive lower ionosphere and propagate efficiently over very long distances with modal attenuation rates of only a few dB per megameter [*Davies, 1990*, pp 367]. Assuming a smooth ionosphere, the mode excitation factors and attenuation rates can be computed as described in *Lehtinen and Inan* [2008, 2009]; *Lehtinen et al.* [2010]. The modal structure is relatively simple at large distances

from the antenna with the bulk of the contributions to the Poynting flux being in the QTM1 and QTM2 modes [Lehtinen *et al.*, 2010]. Any perturbation along this path, in the form of modified ground conductivity, ionospheric electron density, or collision frequency, scatters the incoming fields, potentially creating strong perturbations on the received signal amplitude of a coherent signal at a receiver some distance away from the transmitter.

5.2.1 Simulation domain

We carry out the simulation on a box-shaped domain of size 900 km by 400 km on the ground and 100 km in height (Figure 5.1). We subdivide this domain into a number of horizontal layers (in order to enforce a conformal mesh) with layer thickness approximately proportional to the characteristic length scale in the medium, which we define as the wavelength of a whistler-mode wave propagating parallel to the background magnetic field \mathbf{B}_0 . We truncate the upper part of the domain with an absorbing boundary. The input to the domain is a sum of plane wave sources at one face ($x = -200$ km), extending from the ground up through the ionosphere. The waves propagate in the $+\hat{\mathbf{x}}$ direction, interacting with a scattering region in the ionosphere centered at $(x = 0, y = 0)$, propagating forward and finally terminating at another absorbing boundary placed at $x = 700$ km. To avoid numerical difficulties with terminating very oblique waves and since the bulk of the scattered wave energy is in the forward direction [Lehtinen *et al.*, 2010], we use a periodic boundary condition along the $\hat{\mathbf{y}}$ dimension at $y = -200$ km and $y = 200$ km. The lower boundary is perfectly reflecting. We then mesh the domain using Gmsh [Geuzaine and Remacle, 2009], with average element size again proportional to the characteristic length scale, yielding a domain with approximately 2.5 million tetrahedrons. For the fourth order results in this paper, this translates to 35 points per element for a total of ~ 87 million unknowns per field.

5.2.2 Plasma parameters

The background number density, collision frequency, and magnetic field are as described in Lehtinen *et al.* [2010]. We restrict ourselves to a small number of test

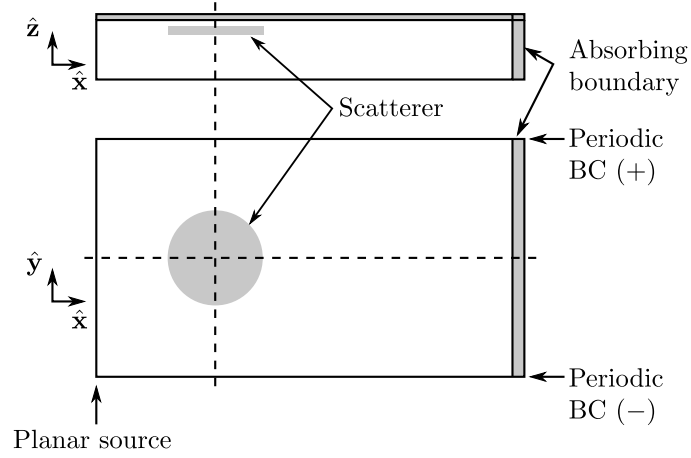


Figure 5.1: Diagram of the simulation domain for the VLF scattering problem, top and side view. The scatterer is located at position $x=0$, $y=0$. Absorbing boundaries are placed at the top and end of the domain. A waveguide mode (or multiple modes) propagating in the $+\hat{x}$ direction are sourced at the input plane indicated. The boundaries in the transverse direction are periodic.

cases. In particular, we restrict ourselves to scattering from ionospheric disturbances created by relatively large lighting discharges. We use the lighting discharge cases described in [Marshall et al. \[2010\]](#) and [Lehtinen et al. \[2010\]](#). For context, in Figure 5.2 we show the ambient electron density and the height-integrated change in the electron density distribution for two of the strong discharges considered in [Marshall et al. \[2010\]](#). The 40V case refers to the density perturbation induced by a large, vertically-oriented lightning discharge. The number 40 refers to the E_{100} parameter. The 10H case refers to a horizontally-oriented inter-cloud, multiple stroke (60 discharges in total) lightning discharge with $E_{100} = 10$ V/m. The background magnetic field is taken to be:

$$\mathbf{B}_0 = B_0(\cos \theta \sin \phi \hat{\mathbf{x}} + \sin \theta \cos \phi \hat{\mathbf{y}} + \cos \phi \hat{\mathbf{z}}), \quad (5.1)$$

where $B_0 = 5 \times 10^{-5}$ T, $\phi = 45^\circ$, and $\theta = 270^\circ$.

5.2.3 Incident source

Since we are computing the scattered field from a narrowband transmitter some large distance away (>1 Mm) from the scatterer we are modeling, it more efficient to use

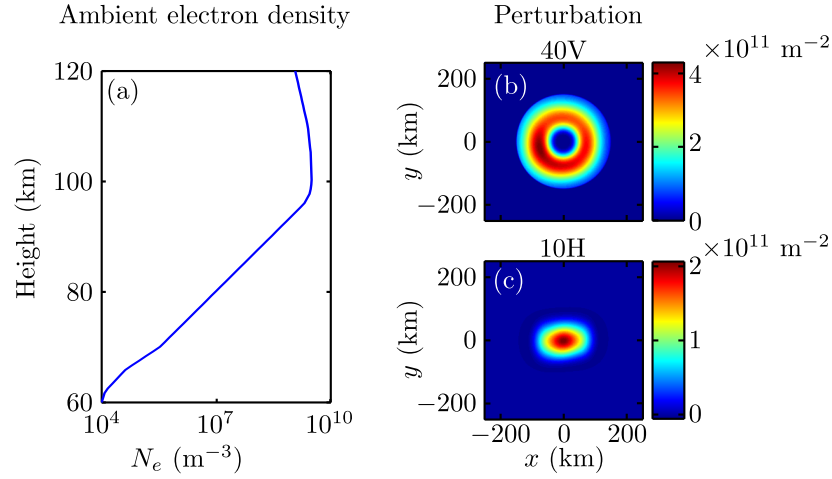


Figure 5.2: (a) The ambient electron density as a function of altitude. (b) Height-integrated electron density perturbation in m^{-2} for a single vertical lightning discharge with $E_{100} = 40$ V/m. (c) Height-integrated electron density perturbation for 60 horizontal discharges with $E_{100} = 10$ V/m.

an approximate method to propagate the fields from a VLF transmitter in the Earth-ionosphere waveguide out to some distance, and then use the field amplitudes at that distance as our incident source to the DG domain, as illustrated in Figure 5.3.

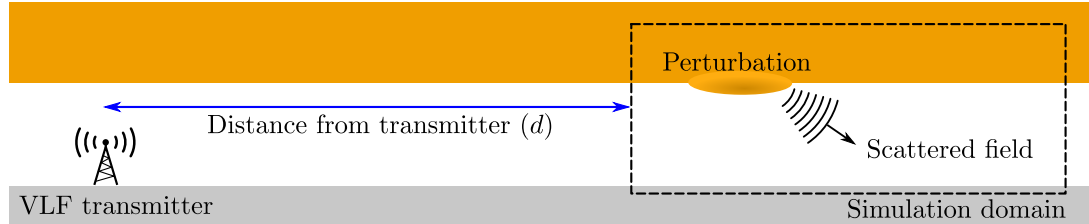


Figure 5.3: Illustration of the combined propagation and scattering model. A VLF transmitter is located some distance from an ionospheric perturbation. These fields are propagated a distance d in the Earth-ionosphere waveguide using a full-wave method to the edge of the simulation domain, providing the incident source terms for the DG simulation. The full 3D scattered fields are computed in the DG domain (denoted with a dashed box).

The propagation method we use is a modal technique described in [Lehtinen et al. \[2010\]](#). The method uses a full-wave technique to compute a set of height gains (that is, the mode structure as a function of altitude) and mode amplitudes at some arbitrary distance from a vertical dipole source. At the scatterer, we assume the cylindrical wavefronts are quasi-planar and then compute a set of equivalent source currents for the given mode (or sum of modes) along a large plane (shown in Figure

5.1) on one side of the domain, where $\hat{\mathbf{n}}$ is the normal to the incident source plane, using the convention that the waves propagate in the same direction as $\hat{\mathbf{n}}$:

$$\begin{aligned}\mathbf{J}_s &= \hat{\mathbf{n}} \times \mathbf{H} \\ \mathbf{M}_s &= -\hat{\mathbf{n}} \times \mathbf{E}\end{aligned}\tag{5.2}$$

As observed by [Lehtinen et al. \[2010\]](#), the dominant contributions at large distances from the VLF source are in the quasi-TM1 (QTM1) and QTM2 modes. For this study, we separately compute the scattered fields for these two pure modes, and we repeat the computation for the full mode sums at selected distances d from the transmitter. We use the full mode sums because the individual waveguide modes can interact to significantly affect the apparent magnitude of the scattered field as observed from a fixed reference point on the ground [[Marshall and Inan, 2010](#)]. Since amplitude and phase measurements are typically observed by at most only a few VLF receivers within the scattered field of a small disturbance, it is important to also consider the behavior of typical systems under these more realistic propagation conditions.

5.3 Results

We apply our model to a number of representative test cases, two pure waveguide modes and mode sums at three distances. We run the simulation until convergence, or approximately 4 ms in physical time.

We first simulate two simple pure-mode cases. For a relatively smooth ionosphere and far from the source, the bulk of the wave power will be in the QTM1 and QTM2 modes, so we consider these two modes separately. We simulate scattering from these pure modes for two large vertical lightning discharges ($E_{100} = 30$ V/m and $E_{100} = 40$ V/m) and two multiple-stroke, horizontal inter-cloud discharges (60 flashes at $E_{100} = 7$ V/m and $E_{100} = 10$ V/m [[Marshall et al., 2008a](#); [Marshall and Inan, 2010](#)]). The vertical discharge creates a flattened ring-shaped disturbance, peaking between 80 km and 90 km altitude, while the horizontal discharges create very localized, oval-shaped disturbances. We source an incident QTM1 or QTM2 mode at one end of the domain and run the simulation until steady-state is reached. Equivalent complex fields are

computed by finding the sine and cosine parts, found by advancing the steady-state simulation results one quarter wavelength in time at the transmitter frequency (24 kHz). We compute a reference simulation without an ionospheric disturbance and use this to compute the scattered amplitude (in dB) and (unwrapped) phase:

$$\Delta A_{\text{scattered}} = 20 \log_{10} \frac{|\mathbf{E}_{\text{perturbed}}|}{|\mathbf{E}_{\text{ref}}|} \quad (5.3a)$$

$$\Delta \phi_{\text{scattered}} = \angle(\mathbf{E}_{\text{ref}}) - \angle(\mathbf{E}_{\text{perturbed}}) \quad (5.3b)$$

The scattered fields for the vertical 40 V/m discharge are shown in Figure 5.4 for QTM1 and QTM2 modes. The results are qualitatively similar to those in [Lehtinen et al. \[2010, Figure 5\]](#) but with a larger amplitude due to the larger magnitude of the disturbance. The peak amplitude perturbation for the QTM1 case is approximately 0.07 dB, with a phase perturbation of 0.5 degrees. The scattered fields for the QTM2 mode are larger in magnitude, on the order of 0.25 dB in amplitude and 1.2 degrees in phase. The smaller response of the QTM1 mode is primarily due to its incidence angle, since grazing incidence does not interact as strongly with the scatterer. Most

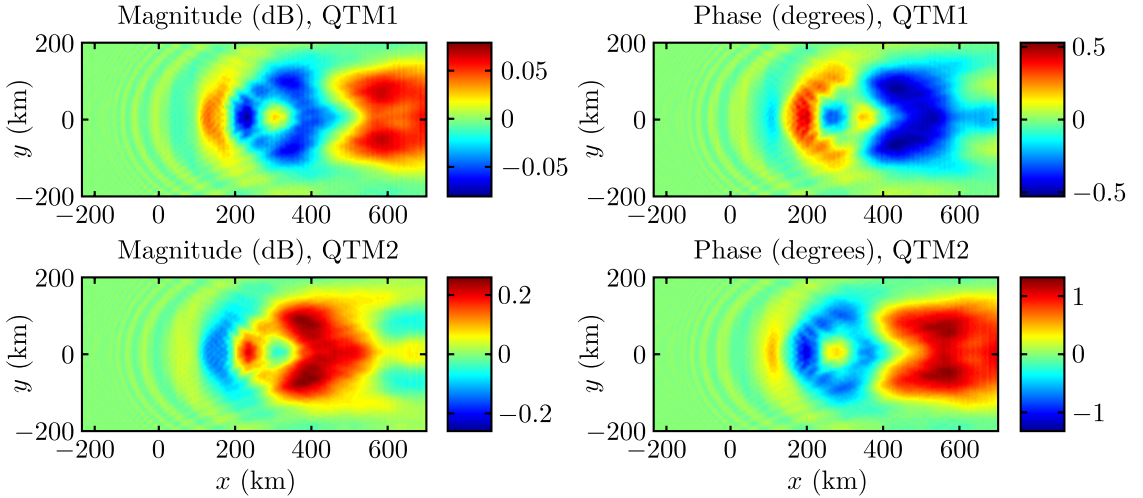


Figure 5.4: Top: Scattered fields for a pure 24 kHz QTM1 mode incident on a vertical ($E_{100} = 40$ V/m) lightning EMP-induced ionospheric disturbance (shown in Figure 5.2). The left panel shows the vertical electric field magnitude response in dB at the ground, while the right is the phase response in degrees (defined in Equation 5.3). **Bottom:** Same, but for QTM2.

striking is the extreme spatial variation in the magnitude of the scattered field, where

the distance from the strongest peak to a null in the scattered field is on the order of 50 km. Also of particular note is the phase response. With some exceptions [Moore *et al.*, 2003, Figure 8], [Johnson *et al.*, 1999, Figure 4], published plots showing early VLF events tend to emphasize the importance of the amplitude response of a received VLF narrowband signal [Johnson and Inan, 2000; Marshall *et al.*, 2008b; Johnson *et al.*, 1999; Marshall *et al.*, 2006; Moore *et al.*, 2003]. Noting, however, that for these simple single-mode cases, the phase response exhibits strong peaks where there is a null in the magnitude response (and vice-versa), the phase response is also an important quantity.

The results for the 60-stroke horizontal 10 V/m discharges are shown in Figure 5.5. The horizontal discharges produce a larger perturbation than the vertical, on the order of 0.3 dB with a maximum phase perturbation of approximately 2 degrees for the QTM1 mode and nearly 1 dB and 6 degrees in phase for the QTM2 mode. We also conducted simulations for a 30 V/m vertical discharge and a multiple-stroke

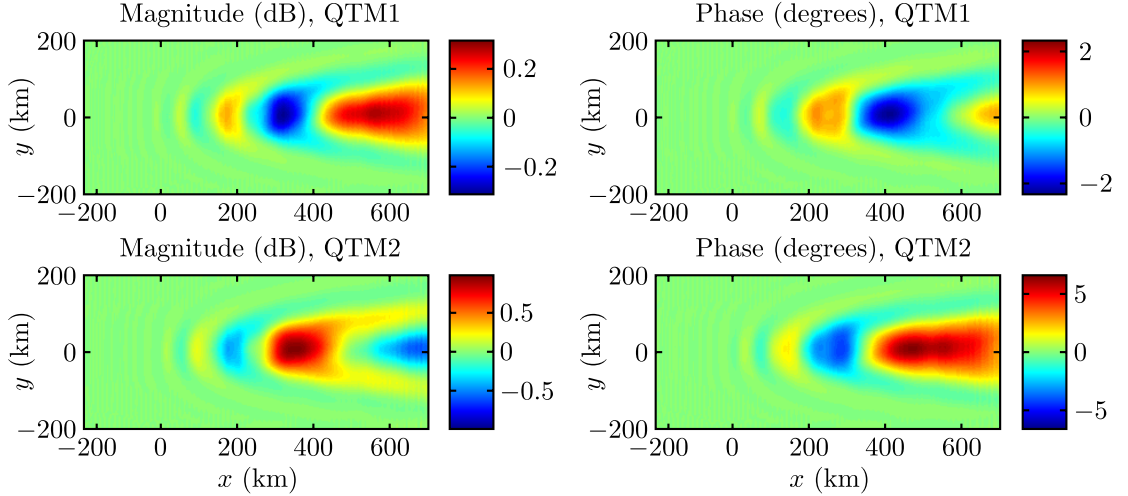


Figure 5.5: Top: Scattered fields for a pure 24 kHz QTM1 mode incident on a horizontal, multi-stroke ($E_{100} = 10 \text{ V/m} \times 60$) lightning EMP-induced ionospheric disturbance (shown in Figure 5.2). The left panel shows the vertical electric field magnitude response in dB at the ground, while the right is the phase response in degrees (defined in Equation 5.3). **Bottom:** Same, but for QTM2.

7 V/m horizontal discharge. The results (not shown) have similar qualitative shapes but smaller perturbation magnitudes, with QTM1 scattering magnitudes on the order of 0.02 dB for the vertical discharge and 0.04 dB for the horizontal discharges. The

QTM2 cases (again, not shown), are also qualitatively quite similar but larger in magnitude, a more measurable 0.25 dB maximum for the vertical 40 V/m discharge (1.5 degrees in phase) and 0.8 dB for the 60-flash horizontal 10 V/m discharge (6 degrees in phase).

For reference, we also show the ambient reference electric field magnitudes as observed on the ground as a function of distance in Figure 5.6 for the cases shown in Figure 5.7. Nulls are visible in the 1000 km case, but the other cases are comparatively smooth.

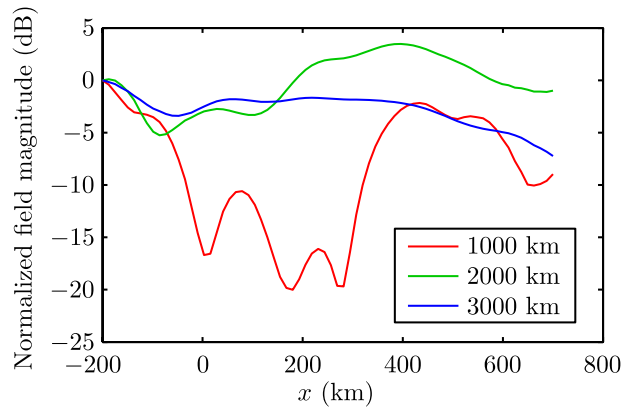


Figure 5.6: Vertical reference electric field magnitude as observed on the ground for the cases shown in Figures 5.7 and 5.8, normalized to their values at $x = -200$ km, plotted for scatterers located at 1000 km, 2000 km, and 3000 km from the transmitter.

For our second set of trials, we use the full mode sums for a VLF transmitter located at some distance from the ionospheric disturbance. As noted by [Marshall and Inan \[2010\]](#), the presence of modal nulls can significantly alter the magnitude of the observed VLF scattered field. As such, we conduct six additional trials for strong scatterers located at fixed distances of 1 Mm, 2 Mm, and 3 Mm from a VLF transmitter. The intent of these trials is to show the shape of the scattered field under more realistic (multi-mode) conditions and to demonstrate that scattered fields with very complicated spatial structure can occur for some configurations, particularly when many waveguide modes are present. The results are summarized in Figures 5.7 and 5.8.

The maximum disturbance measured in these trials is quite large, 5 dB for a lightning discharge 1 Mm away from the VLF transmitter. The horizontal discharges show

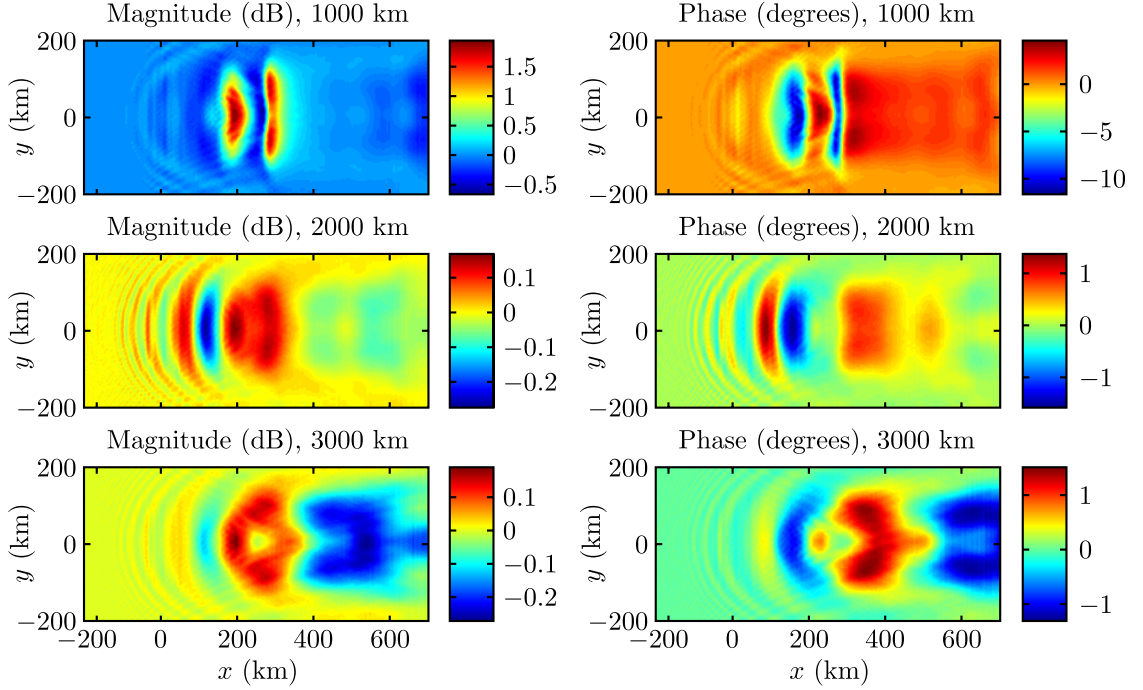


Figure 5.7: Scattered fields as observed on the ground for a disturbance located at a distance from a 24 kHz transmitter. The disturbance is caused by a vertical ($E_{100} = 40$ V/m) lightning EMP-induced ionospheric disturbance (shown in Figure 5.2). The left panels show the vertical electric field magnitude response in dB at the ground, while the right is the phase response in degrees (defined in Equation 5.3). We plot three cases for perturbations located at 1 Mm, 2 Mm, and 3 Mm distance from a vertical VLF antenna.

relatively less transverse variation in the shape of the scattered field, so 2D simulations with cylindrical symmetry could provide reasonable estimates to these scattered fields. The vertical discharges, however, again show strong transverse variation. At larger distances, the QTM1 and QTM2 modes begin to dominate and the scattered fields begin to more closely resemble the single-mode scattering patterns in Figures 5.4 and 5.5. We note that this idealized model assumes perfect uniformity along the path from the transmitter to the lightning-induced plasma perturbation. In reality, the presence of any density irregularities or variations in the ground conductivity along a path excites additional modes. Any such variation from smooth, ambient far-field conditions such as interference nulls, unmodeled scattering from irregularities, or the presence of additional modes may significantly influence the measured magnitude of the scattered field from a single, stationary ground observer. As such, we can say

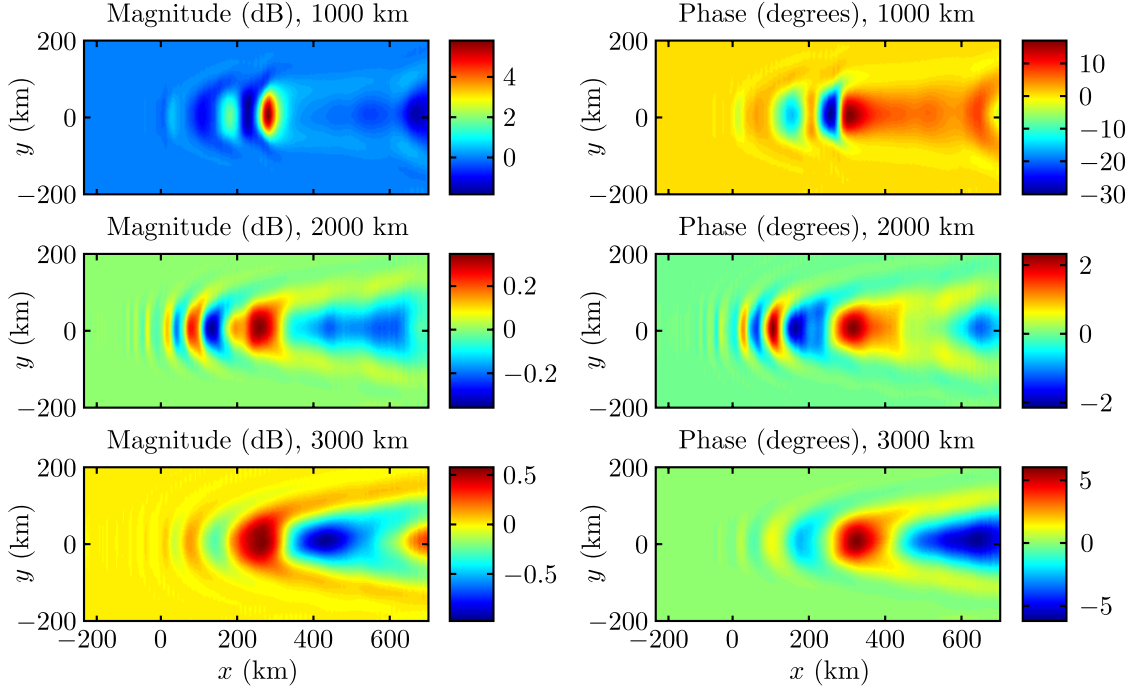


Figure 5.8: Scattered fields as observed on the ground for a disturbance located at a distance from a 24 kHz transmitter. The disturbance is caused by a horizontal, multi-stroke ($E_{100} = 10$ V/m $\times 60$) lightning EMP-induced ionospheric disturbance (shown in Figure 5.2). The left panels show the vertical electric field magnitude response in dB at the ground, while the right is the phase response in degrees (defined in Equation 5.3). We plot three cases for perturbations located at 1 Mm, 2 Mm, and 3 Mm distance from a vertical VLF antenna.

that scattering from lightning-induced electron density perturbations may be consistent with some observations of early VLF events, but it is not possible to make strong inferences about the precise shape and magnitude of the scatterer from a single ground-based measurement.

5.4 Summary

In this chapter, we have modeled scattering of VLF wave energy from lightning EMP-induced ionospheric disturbances over a large 3D volume using our DG framework. As a high order method with geometric flexibility, DG is an ideal tool for direct simulation of wave phenomena in strongly inhomogeneous environments such as the Earth's lower ionosphere. Using a grid with finer sampling only where it is required, e.g., near the VLF reflection height (~ 85 km at night) allows the accurate simulation of domains

much larger than that currently possible using low-order techniques on structured grids. Further, the technique is not limited to small scattered field amplitudes as in Born approximation approaches, thus serving as an independent validation of those methods.

We have shown that the spatial structure of the VLF scattered field from a lightning EMP-induced ionospheric disturbance can be quite complicated, showing rapid variations in the phase and amplitude over a scale of tens to hundreds of kilometers. These simulations demonstrate that scattering from such perturbations may be consistent with observations of so-called early VLF events, with magnitude changes on the order of 0.1 to 1 dB in common conditions and up to 5 dB under some circumstances, consistent with results from the 2D modeling of *Marshall and Inan* [2010]. The bulk of the scattered wave energy is in the direction forward of the perturbation, but the fine-scale structure can be complicated, particularly when signal amplitudes are measured within 700 km of the perturbation. We have further shown that the phase response may also be an important factor to consider when measuring field perturbations from lightning-induced density perturbations.

However, the strong spatial variability in the scattered field and its dependence on ambient conditions further suggests that while some scattering of VLF transmitter signals due to lightning discharges is commonplace, the likelihood of making successful observations of such an event via amplitude and phase measurements of a received narrowband VLF transmitter signal is strongly dependent on the position of the receiver relative to the induced perturbation and the transmitter. This result suggests that the correct approach for analysis of such events is statistical. A large number of similar events caused by similar lightning discharges is necessary to support or invalidate a particular scattering mechanism.

Chapter 6

DG-PIC simulation of spontaneous VLF emissions

This chapter describes DG-PIC simulations of spontaneous emissions in an inhomogeneous magnetized plasma. Specifically, we aim to simulate the spontaneous generation of VLF chorus and hiss waves in the magnetosphere. The intent of this chapter is not to validate or invalidate any particular theory of chorus generation. Rather, we show that DG-PIC is a useful and flexible framework in which to investigate how the properties of the medium affect the generation of VLF waves.

We use this framework to show that for sufficiently unstable electron velocity distributions in an inhomogeneous background magnetic field, we can spontaneously, consistently, and naturally generate rising, quasi-periodic emissions that show good qualitative agreement with chorus waves. Further, our simulation shows that the simulated chorus amplitudes are quite small in the generation region (near the magnetic equator), with amplitudes increasing exponentially as the chorus elements propagate away from the equator, finally reaching saturation some distance away as the trapped energetic particle flux becomes smaller.

We postulate, based on these results, that while nonlinear mechanisms clearly play a role in the generation of chorus elements, consistent, for example, with the results of [Omura *et al.* \[2008\]](#), the primary factor determining whether or not a chorus element can be observed is almost entirely due to simple linear amplification as the wave

propagates away from the equator. This conclusion supports work by [Jordanova et al. \[2010\]](#), who showed that global simulations of anisotropic ring current electron populations produced linear growth rates that were well correlated with the observed distribution of chorus waves.

We begin with a brief discussion of prior work in modeling these phenomena, followed by a description of our 1D model. We then discuss the computational limits of performance of the PIC scheme, which motivates our choice of parameters for the simulation work in the following section.

6.1 Background and prior work

Attempts at simulating spontaneous and triggered emissions from the magnetosphere date back to the 1970s [[Nunn, 1974](#)]. The problem, however, is extremely difficult to solve in the most general case. As such, computational and theoretical approaches usually begin with drastic simplifications in order to make the simulation more feasible. Three basic approaches are common in the literature.

A theoretical approach is to solve for the time evolution of a single particle trajectory in phase space in an inhomogeneous (but smooth) background, subject to forcing by an external wave. We have described this approach in Section 2.3.4. This method provides insight into the processes that may drive an instability and determine the frequency-time characteristics of chorus wave packets. However, it provides no insight into how the external forcing wave forms, how deep a particle trap may become, or how waves with particular frequency-time characteristics may be preferentially selected and amplified. Additionally, the inhomogeneity parameter S proposed by [Omura et al. \[2008\]](#) has an explicit dependence on the perpendicular velocity v_{\perp} of a given particle. It is unclear what effect including a range of perpendicular velocities would have on the underlying theory or the conclusions drawn.

A more direct approach due to [Nunn \[1990\]](#) attempts to evolve an electron distribution function directly in phase space using a first-order finite-volume style method coupled with a drastically-simplified narrowband version of Maxwell's equations. By dropping the displacement current and assuming quasi-monochromaticity of a wave and smoothness of both the wave field and background plasma, the author reduces the

full set of Maxwell's equations to a simple set of equations describing the time evolution of the amplitude and phase of a narrowband wave packet as it progresses through a system. While this method describes the time evolution of a single monochromatic wave as it propagates through an unstable medium, the modeler must a priori impose an input wave frequency and an expected bandwidth before the simulation can even be started. While we acknowledge that the approach is extremely efficient, we do not consider it general enough to simulate spontaneous emissions in an unstable plasma unless one resorts to making a number of ad-hoc assumptions about the background wave environment.

The most direct approach is the PIC method, which self-consistently solves the time evolution of a population of test particles. This method is typically very computationally expensive so authors who use this approach also make some approximations to make the method more feasible. The approach of *Katoh and Omura* [2006], for instance, is to sample only the hot particle distribution using the PIC method while using a cold fluid approach to handle the background, cold plasma. The authors additionally assume strictly parallel propagation along a magnetic field line, which allows the electrostatic component of the electric field (the component parallel to the background magnetic field) to be dropped.

Our approach is most similar to that of *Katoh and Omura* [2006], the major difference being the field solver. We use a high order DG method to evolve the fields, while they use an FDTD method. As discussed in Chapter 4, the DG-PIC scheme has excellent stability and noise properties, and thus we expect at least moderate improvements in the overall quality of the solution for a lower cost (i.e., fewer particles per wavelength). Additionally, we drop the nonlinear convective acceleration term when evolving the cold background plasma. This term is typically small relative to the other terms, and including it would unnecessarily complicate the overall scheme.

6.2 Computational model

The basic computational domain is illustrated in Figure 6.1.

The domain is one-dimensional, where the spatial dimension h is assumed to lie exactly along a magnetic field line intersecting the magnetic equator at some fixed

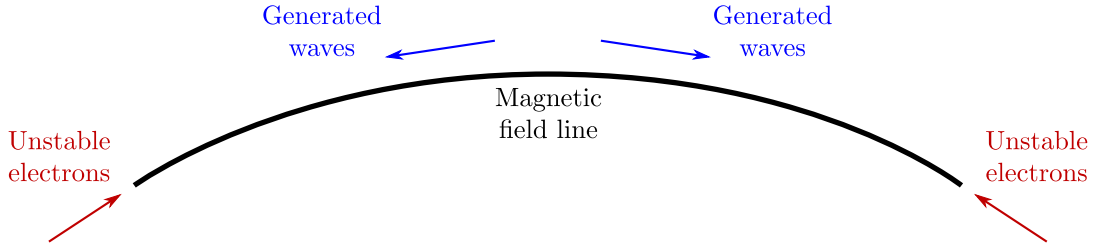


Figure 6.1: Simulation domain for the chorus simulation.

distance L (in Earth radii) from the center of the Earth. Electrons stream in the open boundaries at the left and right while internal, untrapped electrons are free to stream out.

Physically, this one-dimensional simulation can only represent waves that are confined to parallel propagation along a magnetic field line, a situation that only occurs in a so-called *duct*, or magnetic field-aligned density irregularity that can guide waves along its length, much as an optical fiber guides waves in its high-refractive index core [Smith, 1961]. There is abundant observational evidence that the magnetosphere is permeated by density irregularities with a wide range of scale sizes including those supporting ducted VLF propagation [Carpenter *et al.*, 2002; Koons, 1989; Haque *et al.*, 2011].

In order to keep the problem tractable and reduce the computational requirements, we additionally exaggerate the spatial dimension h by some factor α . This exaggeration increases the apparent inhomogeneity of the magnetic field and exaggerates the inhomogeneity parameter S (Equation 2.43) by roughly an equivalent factor, in turn affecting the frequency-time rate at which growth is maximized. The end result is that the simulation time is reduced to approximately one week rather than months.

6.2.1 Mirror force

A one-dimensional simulation implicitly assumes that there is no transverse variation in any of the field quantities or in the background magnetic field. In order to model a trapped particle population, however, it is necessary to add a mirror force back into the system so particles that should be trapped within some range of magnetic

latitudes are actually trapped. This force does not exist in the absence of transverse variation but is necessary to match the real physics of the problem.

The relativistic mirror force is computed as follows [Northrop, 1963], where h is the free spatial coordinate along the field line:

$$F_{\text{mirror}} = -\frac{1}{2} \frac{p_{\perp}^2}{m_0 B} \frac{\partial B}{\partial h} \quad (6.1)$$

This force is added to the Lorentz force (Equation 4.1) acting in a direction parallel to the background magnetic field.

6.2.2 Particle loading

We pre-load the domain by first defining an equatorial bi-Maxwellian particle distribution with three configurable parameters: a loss cone parameter β , a parallel thermal momentum $p_{\text{th},\parallel}$, and a perpendicular thermal momentum $p_{\text{th},\perp}$:

$$f_0(p_{\parallel}, p_{\perp}) = \alpha N_{\text{hot}} \exp\left(-\frac{p_{\parallel}^2}{2p_{\text{th},\parallel}^2}\right) \frac{1}{1-\beta} \left[\exp\left(-\frac{p_{\perp}^2}{2p_{\text{th},\perp}^2}\right) - \exp\left(-\frac{p_{\perp}^2}{2\beta p_{\text{th},\perp}^2}\right) \right]$$

This distribution is illustrated in Figure 6.2 for a typical choice of parameters, plotted in velocity space (see Equation 4.3).

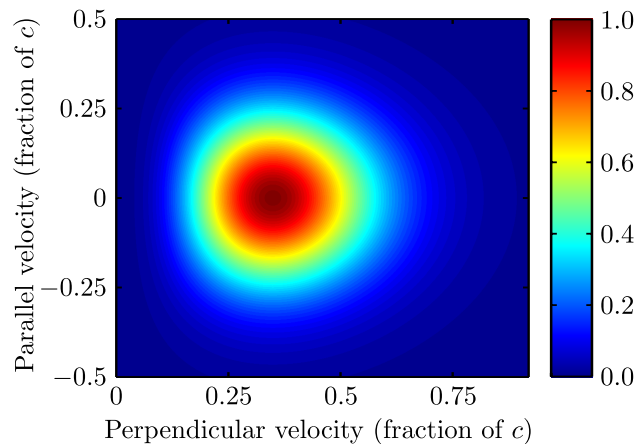


Figure 6.2: Normalized hot particle distribution function for $v_{\text{th},\parallel} = 0.17c$, $v_{\text{th},\perp} = 0.23c$, and $\beta = 0.6$ plotted as a fraction of the speed of light c . High values of β (close to 1.0) give larger depletions at low perpendicular particle velocities. The distribution approaches a plain bi-Maxwellian distribution as $\beta \rightarrow 0$.

To fill the rest of the domain, we assume bounce-adiabatic invariance [[Northrop, 1963](#)] and that the particle population is stably trapped, giving the distribution function as a function of the spatial dimension h :

$$f(p_{\parallel}, p_{\perp}, h) = f_0 \left(\sqrt{p_{\perp}^2 \left(1 - \frac{B_0}{B(h)} \right) + p_{\parallel}^2}, \sqrt{p_{\perp}^2 \frac{B_0}{B(h)}} \right), \quad (6.2)$$

where B_0 is the equatorial magnetic field magnitude and $B(h)$ is the field magnitude as a function of the distance h along the field line.

6.2.3 Particle sampling

The particle-in-cell method requires that we choose a finite number of particles to sample a given continuous distribution f . Choosing the actual sample locations and weights, however, is far from straightforward. Our ultimate goal is to draw from the distribution in such a way that the sample distribution (i.e., PIC particles) converge to the continuous distribution f as the number of samples N grows very large. While there are many sampling techniques that have this property, not all have identical behavior in PIC simulations. For instance, ordered sampling leads to instabilities in PIC simulations due to the presence of nonphysical correlations in the phase space density (either temporal or spatial). Randomized sampling, by contrast, leads to stable schemes, but the resulting noise levels may be excessively high for some applications.

Cold start sampling

Weighted *cold start* sampling uses an ordered grid of sample points, where each particle is weighted by f and normalized by the ratio of PIC particles to real particles. Such a distribution has extremely favorable noise properties but is unfortunately only marginally stable and will eventually thermalize to a random distribution. One way to see this is to treat the system as a system of beams. By Equation 2.40, we can easily verify that almost any multi-beam system is unstable for some range of spatial frequencies. As such, we restrict the use of cold-start sampling to simple linear growth and damping rate calculations, where only a few periods of oscillation are required to determine the rates to fairly high accuracy.

Uniform weighted random sampling

Another strategy is to define a range of velocities over which samples exist and draw from a uniform random distribution within this range. Each sample is then weighted by the particle distribution function f and then normalized by an appropriate ratio such that the particle distribution integrates to the true particle density N_{hot} . This sampling strategy tends to lead to noisy PIC simulations but does not suffer from the multi-beam instability that plagues cold start sampling. As such, it is appropriate for simulating systems for very long time intervals. The disadvantage is that some particles begin with large weights and thus artificially represent very many physical particles while others start with small weights and thus represent many fewer. While this scheme can accurately represent a given distribution at the start of a simulation, as time progresses, the larger weight particles may drift into regions of phase space far from their initial positions. This drift incorrectly biases the distribution towards the heavy particles, leading to excessive numerical noise.

Metropolis sampling

Metropolis-Hastings sampling is a technique to draw samples from a given distribution f . The procedure begins with a sample \mathbf{u}_t representing the current state. A candidate sample \mathbf{u}_c is drawn from another distribution g with the same domain as f and is then either accepted (in which case it becomes the new \mathbf{u}_t) or rejected (in which case the old sample \mathbf{u}_t is re-used). Algorithm 3 given below describes the method in detail.

Metropolis sampling is fast but tends to generate samples with undesirable correlations in velocity space, leading to noisier PIC simulations. If the variance σ^2 is too small, for instance, successive samples will be strongly correlated. If the variance is too large, the rejection rate will be too high, which in the Metropolis algorithm means that many identical samples may be generated in a row. In a PIC simulation, it is computationally wasteful to generate identical samples from a given distribution, since each of the identical particles will follow precisely the same trajectory (within roundoff error) as time progresses.

Algorithm 3 Metropolis sampling a particle distribution f .

```

1:  $\mathbf{u}_t \leftarrow$  Draw an initial sample (anything in the domain of  $f$ )
2: loop
3:    $\mathbf{u}_p \leftarrow$  Draw from normal distribution with mean  $\mathbf{u}_t$  and variance  $\sigma^2$ 
4:    $\alpha \leftarrow \min(1, f(\mathbf{u}_p)/f(\mathbf{u}_t))$ 
5:    $u \leftarrow \text{UniformRandom}(0, 1)$ 
6:   if ( $u < \alpha$ ) then
7:      $\mathbf{u}_t \leftarrow \mathbf{u}_p$ 
8:   else
9:      $\mathbf{u}_t \leftarrow \mathbf{u}_t$ 
10:  end if
11:  Consume sample  $\mathbf{u}_t$ 
12: end loop

```

Rejection sampling

Rejection sampling is another method for sampling from a given distribution f . The method requires that we can draw from another distribution g such that Mg (where M is some scaling constant) dominates f everywhere. The technique is straightforward: given a sample \mathbf{u}_p drawn from such a distribution g , the sample is accepted if $uMg(\mathbf{u}_p) < f(\mathbf{u}_p)$, where u is drawn from the uniform distribution $U(0, 1)$.

Rejection sampling produces high-quality random samples but can be extremely slow and requires that we can find a simple distribution g (and scaling constant M) such that Mg dominates f everywhere. This scaling constant is not always straightforward to compute in practice, but since we are using an explicit expression for f , this is not an issue for our simulations.

In our simulations, we use cold start random sampling to investigate the saturation thresholds for linear growth, since these simulations require very low noise levels. For simulations of nonlinear phenomena over very long time periods, we use rejection sampling. Cold start sampling over such long time intervals would require prohibitively large numbers of particles in order to avoid the multi-beam instability.

6.2.4 Boundary conditions

The domain is terminated on each end by an open boundary. The wave fields are truncated using a pure PML, implemented using Equation 3.39. The aforementioned PML instability (Section 3.2.3) when truncating arbitrary anisotropic media never manifests in one dimension and, further, does not appear for the purely parallel modes (with respect to the background magnetic field) considered in this simulation.

The particle boundaries are handled with a simple inflow-outflow condition. Any particle that travels outside of the domain is discarded. Fresh particles are fed in each end to replenish the lost particles using a first-order advection method, illustrated in Figure 6.3.

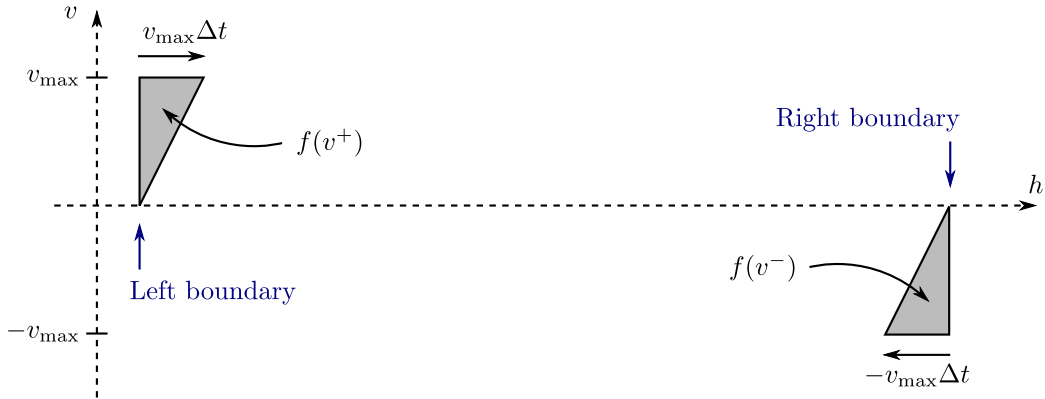


Figure 6.3: Illustration of particle boundary conditions for the chorus simulation. The horizontal axis is the spatial dimension h while the vertical is the particle velocity along h . The outer boundaries of the 1D domain are indicated in blue, where we inject a set of new particles (shaded region) after each timestep Δt . Using first-order advection, all new particles will be contained within the triangular shaded regions indicated. The injected particles are thus found by sampling the positive part of f (on the leftmost boundary) and the negative part of f (on the rightmost) over these shaded regions.

6.3 Saturation and noise

In any real system, a given electromagnetic wave cannot be amplified indefinitely. Not only is the free thermodynamic energy in a given system finite, but there are also other physical processes that tend to limit the energy that can be extracted from a system at a given frequency. While many physical processes can lead to

saturation, in the context of this work, we are primarily concerned with the *linear saturation* amplitude for a uniform, infinite magnetized plasma, which we define as the amplitude where linear cyclotron growth stops for a given whistler mode wave propagating in an infinite, uniform domain.

Direct prediction of the saturation amplitude is exceedingly complicated. Conceptually, we can say that saturation is due an increase in entropy. The wave grows as free energy in the anisotropic distribution function is converted to electromagnetic energy, eventually reaching a point where the available energy for wave growth is completely depleted. However, this simple picture is complicated by the strong wave amplitudes and consequent large perturbations to the electron distribution. The Landau and cyclotron growth rates were originally derived by linearizing the Vlasov equation, that is, by assuming that any perturbation to the electron distribution function f is small. If the original particle distribution anisotropy is sufficiently destroyed, then by definition the perturbation is not small and the assumptions made to derive the growth rate no longer apply. Further, we cannot reasonably expect linearity to hold as the wave amplitude grows arbitrarily large, since large field values should significantly alter the particle distribution from its equilibrium state. Thus, while our intuitive expectation is that wave growth must stop at some amplitude, linear Vlasov theory cannot predict where this termination of growth should happen.

We can gain some insight, however, into the large-amplitude behavior by considering particle traps. Large wave amplitudes effectively split the particles into two non-interacting populations: a population trapped in the wave field and an untrapped population freely gyrating around the background magnetic field, as illustrated in Figure 2.4. The trapped population, while initially ordered with respect to the phase of the driving wave, does, with a timescale on the order of a trapping period (Equation 2.42), become mixed in phase space (i.e., disordered) such that the distribution within the trap effectively thermalizes and the anisotropy is lost. Since the two populations are still isolated from each other, if the trap is deep, then the presence of the trap can also act to suppress linear growth at nearby frequencies. In other words, energy cannot be extracted arbitrarily over the entire available frequency band.

We also note that the number of interacting particles increases along with the

linear growth rate. At infinitesimal linear growth rates, only particles exactly in resonance with the wave, i.e., those for which the parallel velocity equals v_{res} , contribute to wave growth:

$$v_{\text{res}} = \frac{\omega \pm \omega_c}{k_{\parallel}}, \quad (6.3)$$

where the sign \pm is positive for the right-hand circularly polarized whistler mode wave, and ω_c is negative by convention for electrons.

At moderate growth rates, however, particles not exactly in resonance with the wave can also contribute strongly to linear growth. Mathematically, we can interpret this result as the action of the singularity in Equation 2.31. In the small-growth limit, the singularity contributes an arbitrarily large weight to those particles with parallel velocities within a vanishingly small neighborhood about the resonant velocity v_{res} . However, as the linear growth rate becomes larger, the singularity shifts further off the real axis (see Figure 2.3), weighting an ever-larger region of parallel velocities about v_{res} .

Analytically predicting the saturation amplitude from these principles is unfortunately exceedingly difficult, but the basic physical processes at work can nevertheless be described. A newly-trapped population of particles becomes phase-mixed (disordered) on a timescale on the order of one trapping period. This disordered population of particles can no longer coherently exchange energy with the wave and thus cannot contribute to linear growth. This phase mixing is, fundamentally, what causes saturation in a uniform domain. However, the situation is greatly complicated by the fact that the wave is growing and thus the depth of the trap is also increasing. As time progresses, the trap grows larger and an ever-increasing number of particles fall into the wave potential well. Since these newly-trapped particles are not yet entirely phase-mixed, they can still contribute to linear growth for a time (on the order of a trapping period) and linear amplification of the wave continues.

As just discussed, the range of parallel velocities at which particles can significantly contribute to linear growth increases with the growth rate itself. Linear wave growth therefore continues until the “width” of the trap in v_{\parallel} -space eventually grows so large

that it envelops the entire range of parallel velocities that could have contributed to linear growth. It is when this stage is reached that linear growth stops, since all of the particles that could have contributed to wave growth are trapped and phase-mixed. We can therefore conclude that the ultimate achievable depth of the trap (and thus, the saturation amplitude) depends on the linear growth rate itself.

6.3.1 Determination of linear saturation amplitudes

Determining the linear saturation amplitude is essential to obtaining correct results from a chorus PIC simulation. Since chorus simulations must be run for very long periods of time relative to the cyclotron period, quiet start particle distributions cannot be used because the number of parallel velocity samples required to maintain stability over very long time integrations becomes prohibitively expensive. Since we are therefore restricted to only random samples, we must ensure that the artificial noise floor introduced by the particle sampling is far enough below the amplitudes of interest. If the numerical noise level were larger than the saturation amplitude, then we would have no hope of obtaining a physically meaningful solution.

Given how complicated this situation is to treat analytically, we instead turn to direct simulation to determine the saturation amplitude on a uniform domain and use the results as a guide to help choose the appropriate sampling parameters and linear growth rates for a simulation on a non-uniform domain.

We first assume an infinite, uniform domain and a single frequency of interest, which makes the system equivalent to a simple 1D periodic domain, one wavelength in size. On this domain, we sample the unstable distribution function given by Equation 4.15 using quiet start sampling. Quiet start sampling is required because we wish to determine the saturation level in the absence of any confounding noise effects that random sampling might introduce. To eliminate the multi-beam instability over the simulation time of interest, we use a large number of particles in the parallel velocity direction (on the order of 1000) and relatively few in the perpendicular direction.

We then initialize this system with a small amplitude whistler mode wave. The complex characteristics (i.e., the polarization vectors) are determined using the procedure described in Section 2.2.1. Given the characteristics \mathbf{E} and \mathbf{H} associated with

this mode, the initial conditions for the cold plasma state vector \mathbf{x} at the frequency $s = -j\omega$ are found by solving Equation 3.22b for \mathbf{x} :

$$\begin{aligned}\mathbf{x} &= \mathbf{C}^{-1}(\mathbf{J} - \mathbf{D}\mathbf{E}) \\ &= \mathbf{C}^{-1}(\vec{\sigma}(s)\mathbf{E} - \mathbf{D}\mathbf{E})\end{aligned}$$

The simulation is then run through the linear growth phase until saturation (when growth stops). A high-growth example showing these two phases is shown in Figure 6.4. The linear growth phase is clearly visible as an initial period of exponential amplification (a straight line on this log-scale plot). This phase is followed by saturation which is characterized by mildly-damped long-period amplitude oscillations.

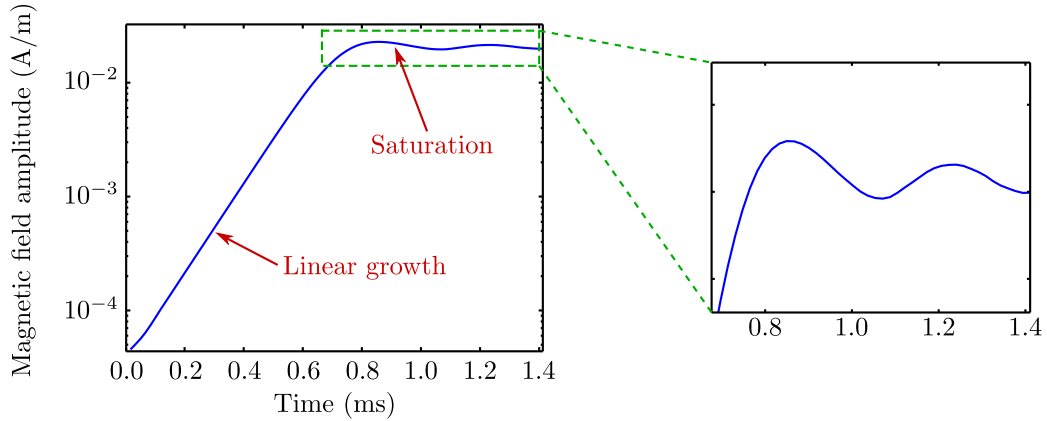


Figure 6.4: Simulation showing linear growth and saturation in an infinite domain. The cold plasma density N is 37 cm^{-3} and the background magnetic field is $0.49 \text{ } \mu\text{T}$ (at the magnetic equator at $L = 4$). The distribution function is given by Equation 4.15 with parameters $v_{\text{th},\parallel} = 0.1c$, $v_{\text{th},\perp} = 0.25c$, $\beta = 0.6$, and $N_{\text{hot}} = 0.1N$.

In the frequency domain, these long-period oscillations appear as sidebands. The bandwidth about the center frequency is approximately equal to the trapping frequency (Equation 2.42) if v_{\perp} is set to some representative perpendicular velocity (e.g., the perpendicular thermal velocity) and B_w is set to the wave amplitude at saturation, B_{sat} :

$$\omega_{\text{sat}} \simeq \sqrt{\frac{kv_{\perp}|q|B_{\text{sat}}}{m}} \quad (6.4)$$

After saturation is reached, the trapped electrons can still oscillate at the trapping frequency ω_{tr} for a few periods, modulating the monochromatic input wave, resulting in sidebands and long-term amplitude oscillations. The oscillations eventually decay as the trapped population becomes phase-mixed and effectively thermalizes.

As discussed in the previous section, the saturation amplitude is sensitive to both the hot particle density and the anisotropy parameter A (Equation 2.34), as shown in Figure 6.5.

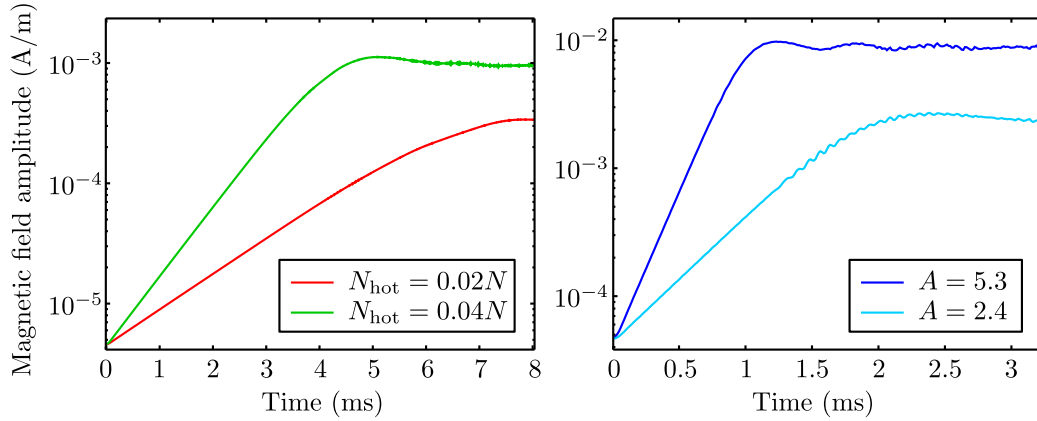


Figure 6.5: Simulations showing linear saturation in an infinite domain. The cold plasma density N is 37 cm^{-3} and the background magnetic field is $0.49 \mu\text{T}$ (at the magnetic equator at $L = 4$). **Left:** Effect of varying the hot particle density N_{hot} as a fraction of the cold density N at a fixed anisotropy of $A = 1.9$. **Right:** Effect of varying the anisotropy at a fixed hot particle density $N_{\text{hot}} = 0.05N$.

6.3.2 Discussion

The results in the previous section suggest that the PIC method, when applied to the chorus growth problem, is only a practical choice for a specific range of parameters, namely cases with very large linear growth rates. Particle noise is the limiting factor for most PIC simulations, particularly in electromagnetic wave-dominated problems such as the one discussed in this chapter. Because the number of PIC particles is finite and much less than the number of actual particles in the system, the noise levels are also larger than in nature. While this problem can be mitigated somewhat by using smoother particle shape functions, doing so only changes the noise by a small multiplicative factor. The largest gains are realized by increasing the number of PIC

particles N_p , but the reduction in noise is unfortunately relatively meager with the noise level going down as $N_p^{-1/2}$.

In the context of this work, this result means that we must choose between either using an enormous number of particles or increasing the linear growth rate. Increasing the number of particles reduces the electromagnetic noise levels, increasing the usable dynamic range of the simulation but at great computational cost. Artificially increasing the growth rate increases the saturation amplitude, which increases the usable dynamic range of the simulation without additional cost.

This tradeoff is the fundamental reason that (to date) successful PIC simulations of chorus growth have all used extremely large linear growth rates. Our technique shares the same limitations, so our simulations also use artificially large anisotropies and particle fluxes far larger than those commonly found in the radiation belts. While the basic physics of the process remains unchanged, the timescales and wave amplitudes are exaggerated from those found in nature.

6.4 Spontaneous emissions

We now turn our attention to simulated spontaneous emissions. We discretize a one-dimensional domain along a magnetic field line within a region near the magnetic equator as in Figure 6.1 and pre-load it with an unstable electron distribution. Our parameters of interest are the cold particle density N , the equatorial hot particle density N_{hot} , the PIC particle density N_p , the inhomogeneity exaggeration factor α , the anisotropy factor A , and the L value. The L value specifies a dipole magnetic field line by the radial distance (in Earth radii) from the center of the Earth to the point where the field line crosses the magnetic equator.

The peak linear growth rate γ is derived from the parameters above and can be specified in natural units (Nepers/s) or in dB/s. For computational tractability, it is necessary to exaggerate two of these quantities beyond their natural values. First, we set the inhomogeneity exaggeration factor $\alpha = 8$ for most simulations. This has the effect of scaling the inhomogeneity factor S (Equation 2.43) by approximately the same factor, which in turn would be expected to increase the frequency-time slew rate of rising emissions by approximately the same factor [Kato and Omura, 2011].

Second, we must exaggerate either the hot particle density or anisotropy factor in order to increase the linear growth rates. In both cases, the effect is to increase the dynamic range of the simulation enough so the features of interest are not swamped by numerical fluctuations introduced by finite particle sampling. We determine the linear saturation amplitudes as described in Section 6.3.1 and use this as a proxy for the behavior of the inhomogeneous system. Heuristically, we have found that good results are obtained if the quiescent noise levels are approximately 20 dB down from the linear saturation amplitude at the magnetic equator.

Increasing the linear growth rate, coupled with the fact that we are not modeling any spreading loss or loss from imperfect guiding along magnetic field-aligned density irregularities (i.e., ducting loss), means that our typical wave amplitudes are also much higher than those encountered in nature.

Even with these approximations in place, the typical computation time for a single simulation is prohibitive, on the order of two weeks on 16 Intel E5530 quad-core CPUs operating at 2.40 GHz. As such, we limit ourselves to a few representative samples of interest.

6.4.1 Results

In Figure 6.6, we show the results of a simulation that produces spontaneous chorus-like rising emissions. The magnetic field inhomogeneity is exaggerated by a factor of eight ($\alpha = 8$), which has the effect of scaling the length and time scales of interest by approximately the same factor. We additionally scale the temporal linear growth rate in order to preserve the unscaled system's spatial linear growth rate, that is, if the unscaled system had a linear growth rate of 100 dB/Mm, then the scaled system should have a linear growth rate of 800 dB/Mm.

For context, we plot the normalized linear growth rate for Figure 6.6 as a function of (scaled) distance along the magnetic field line in Figure 6.7. The maximum linear growth rate at the equator is 6000 dB/s, which in unscaled units is approximately equivalent to 750 dB/s.

The simulations clearly show many of the characteristic features of chorus emissions, including the rising tones and quasi-periodic behavior. Rising emissions are

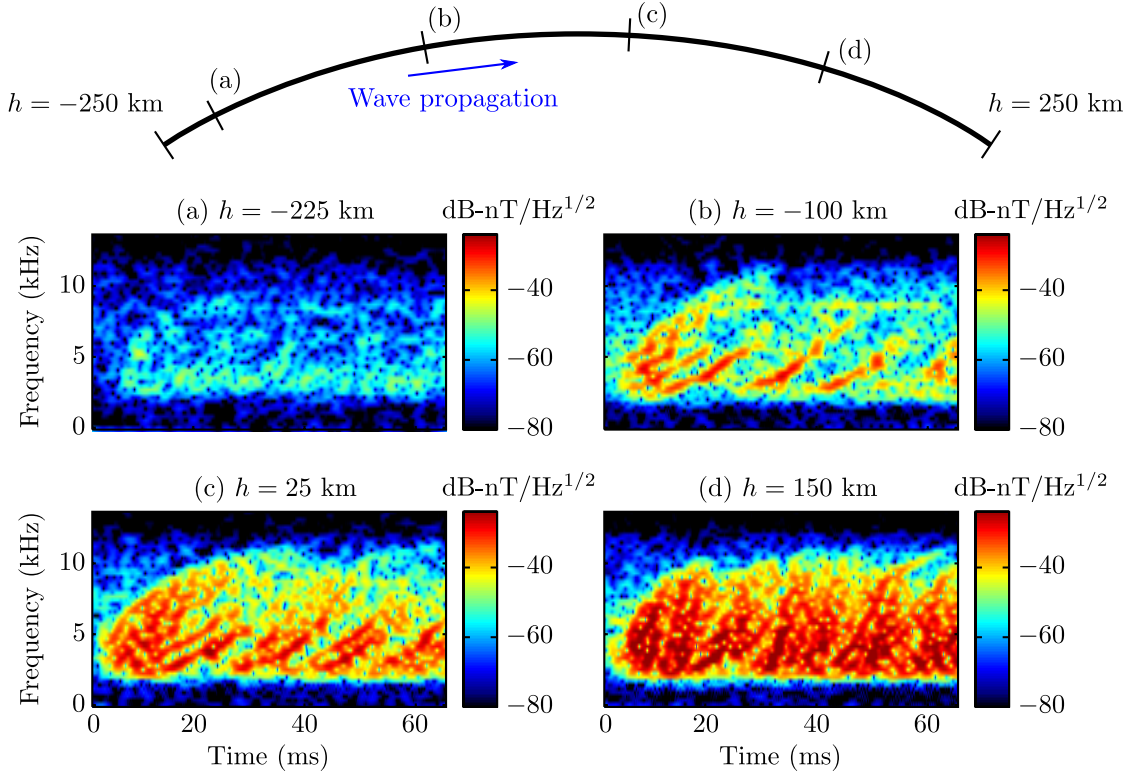


Figure 6.6: Chorus simulation results, showing time-frequency spectrograms at selected distances h along the magnetic field line. The domain is 1D about the magnetic equator at $L = 4$. The magnetic field inhomogeneity (i.e., the spatial dimension) is exaggerated by a factor of 8. The total length of the domain is 500 km (or 4000 km unscaled). The cold plasma density is $37 \text{ electrons cm}^{-3}$. The hot plasma density is $1.85 \text{ electrons cm}^{-3}$ with anisotropy factor of 1.00, corresponding to a linear growth rate of 6000 dB/s (750 dB/s in the unscaled system). The total simulation time extends to approximately 60 ms (500 ms in the unscaled system).

triggered slightly upstream from the equator then amplified as they propagate through the equator and out of the domain. Our simulations have not yet produced falling emissions or hooks. We speculate that these phenomena require smaller growth rates that are difficult to successfully simulate using the PIC technique. We postulate that a direct Vlasov technique would be successful at simulating such phenomena.

The frequency-time sweep rate of the resulting emissions is approximately 201 kHz/s, which in unscaled units would correspond to a sweep rate of 25 kHz/s. Even in unscaled units, this is extremely rapid, but is expected given the high equivalent scaled linear growth rate of 750 dB/s and the correspondingly higher saturation amplitudes [Kato and Omura, 2011]. As demonstrated in Section 6.3.1, the PIC method is

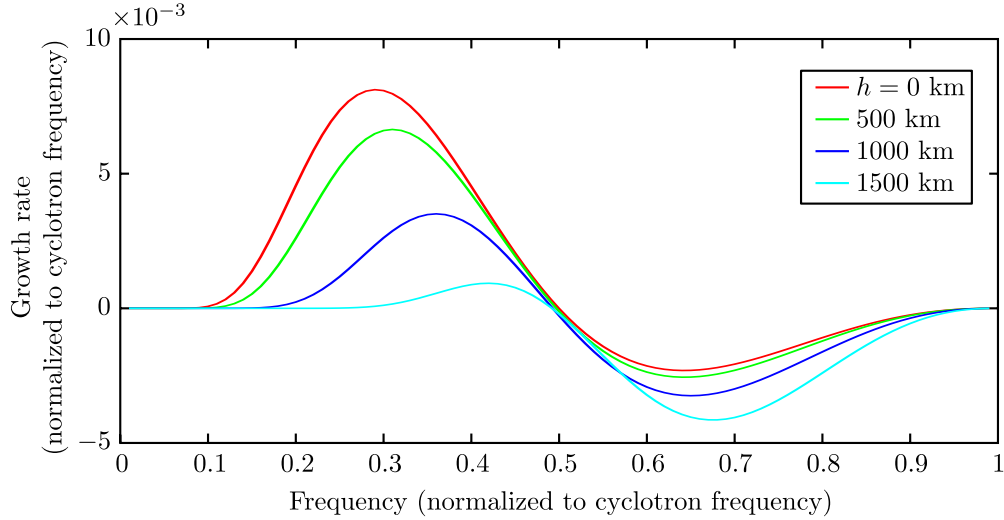


Figure 6.7: Linear growth rates for the scaled system simulated in Figure 6.6 plotted at selected distances from the magnetic equator. The units are normalized to the cyclotron frequency ω_c . The four selected distances are in scaled units, that is, $h = 500$ km represents 4000 km in real space. The linear growth rate drops off rapidly after approximately 1000 km due to the reduction in the trapped particle flux, but is relatively constant over the simulation domain used to generate Figure 6.6.

fundamentally noise-limited to these extremely large linear growth rates.

6.4.2 Spectral properties

According to the theory of [Omura et al. \[2008\]](#), particle trapping is expected to play a dominant role in the development of chorus waves. Trapping is what separates the electron distribution into two discrete populations, enabling the formation of a resonant current whose phase (relative to the wave magnetic field) can vary depending on the local frequency-time slope of the wave field $\partial\omega/\partial t$ and the local slope of the background magnetic field $\partial B_0/\partial h$ (Equation 2.43). This resonant current acts to preferentially amplify background noise with specific frequency-time characteristics, forming chorus seeds which are made visible through subsequent amplification as the seeds propagate away from the generation region.

For such amplification to happen, however, the wave amplitudes must be sufficiently high so that particles can remain trapped by the wave potential well at an amplitude called the *trapping threshold*. In a uniform medium, the trapping threshold

is always met regardless of the wave amplitude, that is, there will always be some population of particles that can be stably trapped by the wave. If there are gradients in the background magnetic field, however, the wave amplitude required to stably trap a population of particles increases [Bell and Inan, 1981]. The trapping threshold is also dependent on the equatorial pitch angle of the particle (the angle between the perpendicular component of the particle’s velocity and the background magnetic field), the wavenormal angle with respect to the magnetic field [Bell, 1984], as well as the local frequency-time characteristics of the wave of interest [Omura *et al.*, 2008].

For this study, it suffices to say that the trapping thresholds are typically lowest near the magnetic equator where the local gradients in the magnetic field are zero and increase sharply away from the equator. This dependence means that the wave amplitudes required to induce nonlinear effects related to trapping are unlikely to be encountered except within a few degrees of the magnetic equator. However, at least some linear amplification is still required to produce wave amplitudes large enough to initially exceed the trapping threshold and trigger nonlinear effects. Further, chorus elements are initially triggered near the equator where the trapping threshold is low but propagate away from the generation region into a region where the required amplitudes are too large to continue trapping particles. These two observations, that trapping requires an initial phase of linear growth and that chorus elements are linearly amplified as they propagate away from the generation region, suggests that there should be a relationship between the spectra of individual chorus elements and the linear growth spectrum.

A sample illustrating this relationship is shown in Figure 6.8, which shows the theoretical equatorial linear growth rate plotted as a function of frequency (left panel) and the observed time-averaged chorus spectra for forward propagating modes (right panel), plotted at three points across a domain 1000 km in size.

At $h = -500$ km (in red), the observed wave spectrum is relatively small in amplitude, diffuse, and unstructured, characteristic of numerical particle noise. As the waves propagate towards the equator ($h = 0$ km), a wide band of wave energy between approximately 0.2 and 0.8 (in normalized frequency) develops, peaking at approximately 0.35. This peak coincides almost exactly with the peak in the linear growth

rate plotted in the left pane. At low frequencies, between the range of 0.2 to 0.35, the observed chorus spectrum and the linear growth rate match quite closely, indicating that linear growth is the predominant process driving wave growth within this frequency band. In the range of 0.35 to 0.80, however, there is a distinct flattening or broadening of the spectrum, caused by nonlinear amplification and generation of rising chorus elements. These elements are initiated near the peak in the linear growth rate (approximately at a normalized frequency of 0.35) and thereafter rapidly increase in frequency.

At the other extreme of the domain ($h = 500$ km), the upper end of the spectrum in the frequency range of 0.5 to 0.8 shows clear evidence of erosion, trending towards closer agreement with the linear growth rate spectrum shown in the left panel. This result suggests that as chorus waves propagate away from the generation region near the equator, linear growth again begins to dominate the amplification process.

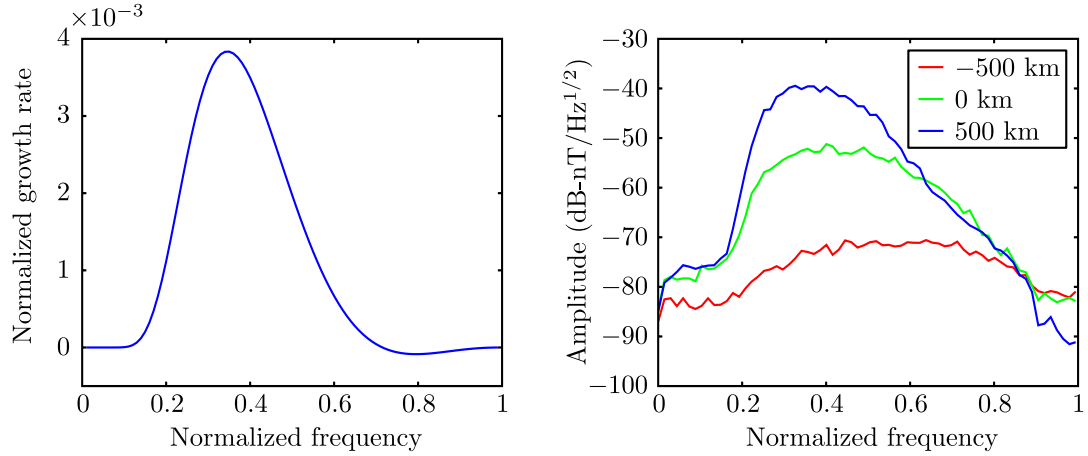


Figure 6.8: Comparison of the equatorial linear growth rate and the observed chorus amplitude spectra at selected locations within a 1000 km domain. The magnetic field is chosen at $L = 4$. The magnetic field inhomogeneity is exaggerated by a factor of 8. The cold plasma density is 37 cm^{-3} . The hot plasma density is 0.2 cm^{-3} with an anisotropy factor of 2.0. **Left:** normalized equatorial linear growth rate (with respect to the cyclotron frequency) plotted as a function of normalized frequency (again with respect to the cyclotron frequency). **Right:** time-averaged spectra for waves propagating in the positive direction, plotted at three positions, at wave initiation (-500 km), at the equator (0 km), and at the other extreme end (500 km).

6.5 Summary and commentary

While not specifically documented here, we conducted other trial simulations for a variety of parameters, some successfully producing coherent, rising emissions and others producing only hiss-like noise. We have made a number of number of interesting observations from these trials.

First, provided that the hot plasma distribution is sufficiently smooth and the linear growth rates are sufficiently large, these simulations consistently and reproducibly generate spontaneous, coherent chorus-like emissions. When the linear growth rates are smaller, the emissions became more unstructured and incoherent, more closely resembling hiss emissions instead of chorus emissions, suggesting that both large linear growth rates and (spatially) smooth hot electron distributions are prerequisites for coherent chorus generation.

While these simulations are practically limited to relatively large growth rates, they nonetheless provide useful insight into the chorus generation process. First, rising chorus emissions seem to be generated predominantly upstream of the magnetic equator. They are subsequently amplified to detectable levels as they approach and pass through the equator. The repetition rate and frequency-time sweep rate both increase as the linear growth rate increases. Further, the linear growth rate can be used to predict many of the gross features of chorus emissions. The frequency of peak linear growth coincides with the peak in the chorus spectrum, and the frequency of peak linear growth is also a fairly reliable predictor of the chorus initiation frequency. These observations suggest that linear growth plays a strong role both in the initial amplification of broadband seed waves to the high amplitudes required to initiate nonlinear growth, and in the subsequent amplification of chorus elements as they propagate away from the equatorial generation region.

While nonlinear growth and triggering clearly plays a role in this process, these results suggest that linear growth rate calculations can serve as an excellent predictor of both the spectra of chorus elements as well as their intensity, sweep rates, and generation locations.

Chapter 7

Summary and Suggestions for Future Work

In this chapter we discuss the limitations of this work and make suggestions for future improvements. We present a number of advances in numerical modeling techniques for nonlinear and linear plasmas, but a number of open modeling and science challenges still remain. Plasma physics remains, and will remain for the foreseeable future, difficult to understand by analytical techniques alone. The sheer variety of phenomena commonly encountered in space and laboratory plasmas, coupled with the pervasive presence of nonlinearities that defy analytical treatments, makes numerical modeling an important component of any investigation into plasma behavior.

We have described the application of a relatively new numerical method, the discontinuous Galerkin (DG) method, to modeling wave propagation and nonlinear growth in strongly inhomogeneous plasmas in the near-Earth space environment. The DG method combines the best features of the finite volume and finite element methods to yield an accurate and local technique for solving hyperbolic partial differential equations. While the DG method is well-known in the fluid dynamics communities, its adoption in other fields has so far been slow. This work aims to demonstrate that applying DG to wave propagation in plasmas is straightforward and leads to highly flexible numerical schemes, largely freeing the modeler from the constraints imposed by lower-order methods on rectilinear grids.

The DG method, however, should not be considered just one technique but a general approach or philosophy for solving partial differential equations. The DG cold plasma and DG-PIC hybrid methods described here are by no means unique. We now discuss some ways in which the technique could be extended or adapted.

7.1 Adaptivity

The scheme described here is static. The modeler must decide in advance, based on some physical intuition or prior simulations, how many elements per unit space are required to obtain an accurate solution. While the DG method greatly simplifies this process by providing a variable order of accuracy, the modeler must nonetheless decide on a set of parameters and make some educated decision about how much accuracy is necessary for the given problem.

Adaptivity, either grid-based adaptivity or adaptivity in the polynomial basis, simplifies this process. The modeler, instead of deciding on a grid, may decide on an acceptable relative error tolerance, and the scheme adaptively adds elements or increases the polynomial order where required to meet that tolerance. Parallelizing such a scheme can be quite difficult. First, every data structure must be dynamic, which is particularly difficult to handle efficiently for implicit time-stepping schemes since the matrices must be rebuilt and refactored frequently, potentially every timestep. Second, without some means to balance the load between different computational nodes, an adaptive scheme may very quickly become unbalanced, wasting resources.

7.2 DG Basis

The DG basis used here, the interpolating polynomial basis, is not the only possible choice. While it leads to intuitive, efficient schemes, it nonetheless imposes a specific form that the solutions are expected to take. With prior knowledge about the problem at hand, however, a more intelligent basis could be chosen. A so-called *plane-wave* basis is one such example. Instead of approximating the solution on each element using interpolating polynomials, we could instead approximate the solution using a family of plane waves of varying wavelength and direction. For fixed-frequency

radio wave propagation at VLF in the ionosphere, such an approach could entirely remove the spatial sampling restriction. For instance, if a section of space contained ten wavelengths of a wave at a given frequency, we would, at minimum require 20 unknowns per dimension per wavelength in order to even sample the solution, without even considering convergence or accuracy. With a plane wave basis, on the other hand, we would only require a direction, a wavelength, and an amplitude.

The disadvantage of such an approach is that a plane wave basis does not span a continuous solution space very efficiently, meaning that many plane waves may be required to adequately approximate a solution. Because of this requirement, efficient schemes using a plane wave basis typically require some way to adaptively choose the best basis to represent the solution at a given time and given point in space. The question of how to efficiently select a subset of such a family of plane waves is an entire dissertation topic in its own right.

7.3 Higher-order fluid moments

A major assumption in the treatment of waves in plasmas is that a plasma consists of only a cold and hot component. This assumption is typically adequate for radio propagation in magnetospheric plasmas but fails when plasma heating or coupling between the two populations becomes significant. A multi-moment approach (Section 2.1.5) can be used to handle temperature effects, but we have not attempted to derive the DG scheme here. Including temperature effects introduces additional spatial derivatives and a nonlinear flux term, requiring a reformulation of the numerical fluxes as well as minor changes to the scheme itself. Additionally, the question of how to appropriately handle nonlinear fluxes in a high-order DG scheme is still somewhat of an open research topic. Many of the strong results that exist for finite volume schemes with nonlinear fluxes do not, unfortunately, carry over directly to DG.

7.4 PIC

The particle-in-cell method is not particularly well-suited to simulating very low-amplitude, weakly growing waves. Simulating such long-time phenomena requires

random particle samples in order to avoid a non-physical multi-beam instability. Unfortunately, randomly-sampled PIC schemes have very high artificial noise levels, which can easily swamp quiet phenomena of interest. While PIC is a good choice for modeling spontaneous emissions, the high noise levels spoil attempts to model triggered emissions, which are narrowband chorus-like emissions initiated by a monochromatic pump wave. Hybrid Vlasov or direct Vlasov simulation are the only reasonable choices to simulate such quiet phenomena over very long timescales. Unfortunately, direct discretization of even 1D domains is often prohibitively expensive unless adaptive techniques are used or other simplifying assumptions are made (e.g., reducing Maxwell's equations to a set of narrowband wave equations). Researchers who do attempt direct discretization of the Vlasov equation typically must use first-order adaptive finite volume methods. There is no fundamental reason, however, that the DG method could not be used to solve the conservative form (Equation 2.12) of the Vlasov equation directly. Using the DG method directly on the 6D Vlasov equation seems an interesting and fertile area for research. However, many difficult problems remain: how to efficiently discretize a 6D space, how to choose a suitable basis, how to handle adaptivity, and how to derive an efficient numerical flux for the Vlasov equation.

Appendix A

Solution to the inverse cold plasma dispersion relation

Using the procedure outlined in Section 3.2.1, we rewrite the frequency domain description relating the cold plasma current \mathbf{J}_i to the electric field \mathbf{E} for the i^{th} plasma species as a first-order ODE:

$$\begin{aligned}\frac{d\mathbf{x}_i}{dt} &= \mathbf{A}_i\mathbf{x}_i + \mathbf{B}_i\mathbf{E} \\ \mathbf{J}_i &= \mathbf{C}_i\mathbf{x}_i + \mathbf{D}_i\mathbf{E}\end{aligned}$$

For the sake of example, we will restrict ourselves to a two-component plasma characterized by conductivity tensors $\vec{\sigma}_1(s)$ and $\vec{\sigma}_2(s)$, where the subscript denotes the plasma species, that is, realization generates the matrices $(\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1, \mathbf{D}_1)$ from the plasma conductivity tensor $\vec{\sigma}_1(s)$ for species 1 and $(\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_2, \mathbf{D}_2)$ from $\vec{\sigma}_2(s)$. The additional vectors \mathbf{x}_1 and \mathbf{x}_2 represent the internal plasma states for each species. Physical meaning can be ascribed to these state vectors by recalling that the plasma current \mathbf{J}_i for a species i is equal to $\mathbf{C}_i\mathbf{x}_i + \mathbf{D}_i\mathbf{E}$.

Making the substitutions $\partial/\partial k \rightarrow jk$ and $\partial/\partial\omega \rightarrow -j\omega$, the governing equations

for a cold plasma are:

$$\begin{aligned}
 j\mathbf{k} \times \mathbf{H} - (\mathbf{C}_1 \mathbf{x}_1 + \mathbf{D}_1 \mathbf{E}) - (\mathbf{C}_2 \mathbf{x}_2 + \mathbf{D}_2 \mathbf{E}) &= -j\omega\epsilon_0 \mathbf{E} \\
 j\mathbf{k} \times \mathbf{E} &= j\omega\mu_0 \mathbf{H} \\
 \mathbf{A}_1 \mathbf{x}_1 + \mathbf{B}_1 \mathbf{E} &= -j\omega \mathbf{x}_1 \\
 \mathbf{A}_2 \mathbf{x}_2 + \mathbf{B}_2 \mathbf{E} &= -j\omega \mathbf{x}_2
 \end{aligned}$$

The extension to any number of species follows straightforwardly. Rewriting the cross product with \mathbf{k} as a matrix-vector product as in Section 2.2.1, the system can be rewritten as a block matrix equation:

$$\begin{bmatrix}
 -(\mathbf{D}_1 + \mathbf{D}_2) & jk\hat{\mathbf{K}} & -\mathbf{C}_1 & -\mathbf{C}_2 \\
 jk\hat{\mathbf{K}} & 0 & 0 & 0 \\
 \mathbf{B}_1 & 0 & \mathbf{A}_1 & 0 \\
 \mathbf{B}_2 & 0 & 0 & \mathbf{A}_2
 \end{bmatrix}
 \begin{bmatrix}
 \mathbf{E} \\
 \mathbf{H} \\
 \mathbf{x}_1 \\
 \mathbf{x}_2
 \end{bmatrix}
 = \omega
 \begin{bmatrix}
 -j\epsilon_0 \mathbf{I} & 0 & 0 & 0 \\
 0 & j\mu_0 \mathbf{I} & 0 & 0 \\
 0 & 0 & -j\mathbf{I} & 0 \\
 0 & 0 & 0 & -j\mathbf{I}
 \end{bmatrix}
 \begin{bmatrix}
 \mathbf{E} \\
 \mathbf{H} \\
 \mathbf{x}_1 \\
 \mathbf{x}_2
 \end{bmatrix}$$

The task is to find the nontrivial ω and the associated vectors \mathbf{E} , \mathbf{H} , \mathbf{x}_1 , and \mathbf{x}_2 for which this equation holds. This is a generalized eigenvalue problem of the form:

$$\mathbf{A}\mathbf{v} = \omega\mathbf{B}\mathbf{v}$$

However, it is simpler (and more numerically robust) to solve if we use the fact that \mathbf{B} is diagonal and nonsingular and thus simply inverted, yielding the simple eigenvalue problem:

$$(\mathbf{B}^{-1}\mathbf{A})\mathbf{v} = \omega\mathbf{v}$$

Numerically solving this system with an eigenvalue solver, then, solves the dispersion relation simultaneously for all ω in terms of a given k . As before, an added advantage is that we also obtain the characteristic vectors \mathbf{E} and \mathbf{H} associated with these modes.

Appendix B

Derivation of the upwind flux

B.1 Two-characteristic derivation

We consider a linear system of the form:

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

Applying the product rule, we rewrite this system as:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{A} \cdot (\nabla \mathbf{u}) = \mathbf{g}$$

We drop the additional terms \mathbf{g} since they do not directly affect the analysis of the characteristics. Explicitly:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{A}_x \frac{\partial \mathbf{u}}{\partial x} + \mathbf{A}_y \frac{\partial \mathbf{u}}{\partial y} + \mathbf{A}_z \frac{\partial \mathbf{u}}{\partial z} = 0$$

For convenience, we reduce the system to one dimension by projecting along a direction $\hat{\mathbf{n}}$:

$$\frac{\partial \mathbf{u}}{\partial t} + (\hat{\mathbf{n}} \cdot \mathbf{A}) \frac{\partial \mathbf{u}}{\partial n} = 0$$

Or explicitly:

$$\frac{\partial \mathbf{u}}{\partial t} + (n_x \mathbf{A}_x + n_y \mathbf{A}_y + n_z \mathbf{A}_z) \frac{\partial \mathbf{u}}{\partial n} = 0$$

Defining $\hat{\mathbf{A}} = \hat{\mathbf{n}} \cdot \mathbf{A}$, we have:

$$\frac{\partial \mathbf{u}}{\partial t} + \hat{\mathbf{A}} \frac{\partial \mathbf{u}}{\partial n} = 0$$

The characteristics of the system along a direction $\hat{\mathbf{n}}$ are then just the eigenvalues and eigenvectors of the matrix $\hat{\mathbf{A}}$: (λ_1, λ_2) and $(\mathbf{v}_1, \mathbf{v}_2)$.

The upwind flux is derived by solving a Riemann problem at the boundary between two elements. The Riemann problem is a type of initial value problem that begins with an initial discontinuity and then asks for the solution at some $t = t_0$ later. This is usually illustrated graphically in the x - t plane as shown in Figure B.1. The solution

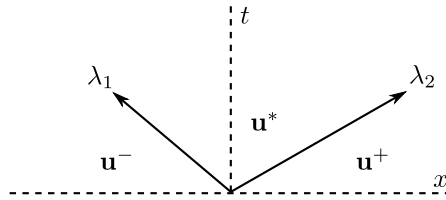


Figure B.1: Graphical illustration of the two-characteristic Riemann problem. The initial conditions are discontinuous at $x=0$ and $t=0$. The initial value left of the discontinuity is \mathbf{u}^- and right of the discontinuity is \mathbf{u}^+ . At $t > 0$, two discontinuities will propagate to the left and the right at the characteristic velocities λ_1 and λ_2 .

to the Riemann problem is found by solving:

$$\begin{aligned} \lambda_1(\mathbf{u}^* - \mathbf{u}^-) &= \hat{\mathbf{n}} \cdot \mathbf{f}^* - \hat{\mathbf{n}} \cdot \mathbf{f}^- \\ \lambda_2(\mathbf{u}^+ - \mathbf{u}^*) &= \hat{\mathbf{n}} \cdot \mathbf{f}^+ - \hat{\mathbf{n}} \cdot \mathbf{f}^* \end{aligned}$$

If the equations are still hyperbolic, the intermediate flux $\hat{\mathbf{n}} \cdot \mathbf{f}^*$ is a good estimate of the value at the interface between two elements. Solving the first equation for $\hat{\mathbf{n}} \cdot \mathbf{f}^*$, we have:

$$\hat{\mathbf{n}} \cdot \mathbf{f}^* = \hat{\mathbf{n}} \cdot \mathbf{f}^- - \lambda_1(u^- - u^*)$$

Suppose we draw a horizontal line in Figure B.1 at some time $t = t_0$. Starting from the left, we are in a region of \mathbf{u}^- . Progressing rightward, there is a jump in the solution from \mathbf{u}^- to \mathbf{u}^* when we encounter the mode traveling at velocity λ_1 . Progressing further rightward, there is another jump as the solution changes from \mathbf{u}^* to \mathbf{u}^+ when we encounter the second characteristic traveling at velocity λ_2 .

The Riemann problem can be solved by observing that the jump magnitudes are proportional to \mathbf{v}_i , the eigenvector associated with the eigenvalue λ_i . Thus, we need to find the unknown jump magnitudes (α_1, α_2) by solving:

$$(\mathbf{u}^+ - \mathbf{u}^*) + (\mathbf{u}^* - \mathbf{u}^-) = \alpha_1 \mathbf{v}_1^- + \alpha_2 \mathbf{v}_2^+$$

Or, since the sum collapses:

$$\mathbf{u}^+ - \mathbf{u}^- = \alpha_1 \mathbf{v}_1^- + \alpha_2 \mathbf{v}_2^+$$

The superscripts on \mathbf{v} reflect the fact that the characteristics themselves may be different on either side of the boundary. The unknown intermediate field \mathbf{u}^* is found by starting at the left and summing rightward over each of the jumps until we reach the intermediate solution \mathbf{u}^* :

$$\mathbf{u}^* = \mathbf{u}^- + \alpha_1 \mathbf{v}_1^-$$

B.2 Solution for the Maxwell system

For the Maxwell system, if $\mathbf{u} = (E_x, E_y, E_z, H_x, H_y, H_z)$, then the matrix $\hat{\mathbf{A}}$ is:

$$\hat{\mathbf{A}} = \begin{bmatrix} 0 & 0 & 0 & 0 & n_z/\epsilon & -n_y/\epsilon \\ 0 & 0 & 0 & -n_z/\epsilon & 0 & n_x/\epsilon \\ 0 & 0 & 0 & n_y/\epsilon & -n_x/\epsilon & 0 \\ 0 & -n_z/\mu & n_y/\mu & 0 & 0 & 0 \\ n_z/\mu & 0 & -n_x/\mu & 0 & 0 & 0 \\ -n_y/\mu & n_x/\mu & 0 & 0 & 0 & 0 \end{bmatrix} \quad (\text{B.1})$$

The nonzero eigenvalues $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ of \mathbf{A} are $(-c^-, -c^-, c^+, c^+)$, where c is the speed of light in the medium, $(\mu\epsilon)^{-1/2}$. We have sorted the eigenvalues into left-propagating modes to the left of the interface (superscript “−”) and right-propagating modes to the right of the interface (superscript “+”), where in general we assume that $|c^-| \neq |c^+|$. The eigenvectors (with an arbitrary scale factor) associated with these eigenvalues are:

$$\begin{aligned}\mathbf{v}_1^- &= (n_x n_y Z^-, -(n_x^2 + n_z^2) Z^-, n_y n_z Z^-, -n_z, 0, n_x) \\ \mathbf{v}_2^- &= (-n_x n_z Z^-, -n_y n_z Z^-, (n_x^2 + n_y^2) Z^-, -n_y, n_x, 0) \\ \mathbf{v}_3^+ &= (-n_x n_y Z^+, (n_x^2 + n_z^2) Z^+, -n_y n_z Z^+, -n_z, 0, n_x) \\ \mathbf{v}_4^+ &= (n_x n_z Z^+, n_y n_z Z^+, -(n_x^2 + n_y^2) Z^+, -n_y, n_x, 0)\end{aligned}$$

The eigenvectors can be clearly interpreted as the two polarizations of waves propagating in the $+\hat{\mathbf{n}}$ direction and two in the $-\hat{\mathbf{n}}$ direction. Proceeding as in the previous section, we solve the following system to find the unknown jump magnitudes α_i :

$$\mathbf{u}^+ - \mathbf{u}^- = \alpha_1 \mathbf{v}_1^- + \alpha_2 \mathbf{v}_2^- + \alpha_3 \mathbf{v}_3^+ + \alpha_4 \mathbf{v}_4^+$$

The intermediate values \mathbf{u}^α , \mathbf{u}^β , and \mathbf{u}^* are found as before by summing over the weighted characteristics from left to right or from right to left. The jumps and fluxes are as before, but now with four jumps:

$$\begin{aligned}-c^-(\mathbf{u}^\alpha - \mathbf{u}^-) &= \hat{\mathbf{n}} \cdot (\mathbf{f}^\alpha - \mathbf{f}^-) \\ -c^-(\mathbf{u}^* - \mathbf{u}^\alpha) &= \hat{\mathbf{n}} \cdot (\mathbf{f}^* - \mathbf{f}^\alpha) \\ c^+(\mathbf{u}^\beta - \mathbf{u}^*) &= \hat{\mathbf{n}} \cdot (\mathbf{f}^\beta - \mathbf{f}^*) \\ c^+(\mathbf{u}^+ - \mathbf{u}^\beta) &= \hat{\mathbf{n}} \cdot (\mathbf{f}^+ - \mathbf{f}^\beta)\end{aligned}$$

Combining all this information, we can solve for $\hat{\mathbf{n}} \cdot \mathbf{f}^*$. After a great deal of manipulation, we finally arrive at our solution for the upwind flux along the normal, which we

have written in the form required by the strong form of the DG scheme, $\hat{\mathbf{n}} \cdot (\mathbf{f} - \mathbf{f}^*)$:

$$\hat{\mathbf{n}} \cdot (\mathbf{f}_E - \mathbf{f}_E^*) = \frac{1}{\epsilon^-(Z^+ + Z^-)} (\hat{\mathbf{n}} (\hat{\mathbf{n}} \cdot [\mathbf{E}]) - [\mathbf{E}] - Z^+ \hat{\mathbf{n}} \times [\mathbf{H}]) \quad (\text{B.2})$$

$$\hat{\mathbf{n}} \cdot (\mathbf{f}_H - \mathbf{f}_H^*) = \frac{1}{\mu^-(Y^+ + Y^-)} (\hat{\mathbf{n}} (\hat{\mathbf{n}} \cdot [\mathbf{H}]) - [\mathbf{H}] + Y^+ \hat{\mathbf{n}} \times [\mathbf{E}]) \quad (\text{B.3})$$

The bracketed notation $[\mathbf{u}]$ denotes the field difference, $[\mathbf{u}] = \mathbf{u}^- - \mathbf{u}^+$. Note that an unsuperscripted \mathbf{f} is the flux interior to the element.

B.3 Solution for the augmented Maxwell system

For the Maxwell system augmented with the hyperbolic divergence cleaning field (Equation 4.7), if $\mathbf{u} = (E_x, E_y, E_z, H_x, H_y, H_z, \phi)$, then the matrix $\hat{\mathbf{A}}$ is:

$$\hat{\mathbf{A}} = \begin{bmatrix} 0 & 0 & 0 & 0 & n_z/\epsilon & -n_y/\epsilon & \chi n_x \\ 0 & 0 & 0 & -n_z/\epsilon & 0 & n_x/\epsilon & \chi n_y \\ 0 & 0 & 0 & n_y/\epsilon & -n_x/\epsilon & 0 & \chi n_z \\ 0 & -n_z/\mu & n_y/\mu & 0 & 0 & 0 & 0 \\ n_z/\mu & 0 & -n_x/\mu & 0 & 0 & 0 & 0 \\ -n_y/\mu & n_x/\mu & 0 & 0 & 0 & 0 & 0 \\ \chi n_x & \chi n_y & \chi n_z & 0 & 0 & 0 & 0 \end{bmatrix} \quad (\text{B.4})$$

The eigenvalues $(\lambda_1, \dots, \lambda_7)$ of $\hat{\mathbf{A}}$ are $(-\chi, -c^-, -c^-, 0, c^+, c^+, \chi)$, where χ is the speed of the divergence cleaning mode. We have sorted the eigenvalues into left-propagating modes to the left of the interface (superscript “−”) and right-propagating modes to the right of the interface (superscript “+”), where in general we assume that $|c^-| \neq |c^+|$ but χ has the same value everywhere. The eigenvectors (with an arbitrary

scale factor) associated with these eigenvalues are:

$$\begin{aligned}
\mathbf{v}_1^- &= (-n_x, -n_y, -n_z, 0, 0, 0, 1) \\
\mathbf{v}_2^- &= (n_x n_y Z^-, -(n_x^2 + n_z^2) Z^-, n_y n_z Z^-, -n_z, 0, n_x, 0) \\
\mathbf{v}_3^- &= (-n_x n_z Z^-, -n_y n_z Z^-, (n_x^2 + n_y^2) Z^-, -n_y, n_x, 0, 0) \\
\mathbf{v}_4 &= (0, 0, 0, n_x, n_y, n_z, 0) \\
\mathbf{v}_5^+ &= (-n_x n_y Z^+, (n_x^2 + n_z^2) Z^+, -n_y n_z Z^+, -n_z, 0, n_x, 0) \\
\mathbf{v}_6^+ &= (n_x n_z Z^+, n_y n_z Z^+, -(n_x^2 + n_y^2) Z^+, -n_y, n_x, 0, 0) \\
\mathbf{v}_7^+ &= (n_x, n_y, n_z, 0, 0, 0, 1)
\end{aligned}$$

We continue as in Appendix B.2, yielding the upwind fluxes:

$$\hat{\mathbf{n}} \cdot (\mathbf{f}_E - \mathbf{f}_E^*) = \frac{1}{\epsilon^-(Z^+ + Z^-)} (\hat{\mathbf{n}} (\hat{\mathbf{n}} \cdot [\mathbf{E}]) - [\mathbf{E}] - Z^+ \hat{\mathbf{n}} \times [\mathbf{H}]) + \quad (\text{B.5})$$

$$\frac{1}{2} \chi \hat{\mathbf{n}} ([\phi] - \hat{\mathbf{n}} \cdot [\mathbf{E}])$$

$$\hat{\mathbf{n}} \cdot (\mathbf{f}_H - \mathbf{f}_H^*) = \frac{1}{\mu^-(Y^+ + Y^-)} (\hat{\mathbf{n}} (\hat{\mathbf{n}} \cdot [\mathbf{H}]) - [\mathbf{H}] + Y^+ \hat{\mathbf{n}} \times [\mathbf{E}]) \quad (\text{B.6})$$

$$\hat{\mathbf{n}} \cdot (\mathbf{f}_\phi - \mathbf{f}_\phi^*) = \frac{1}{2} \chi (\hat{\mathbf{n}} \cdot [\mathbf{E}] - [\phi]) \quad (\text{B.7})$$

Note that an unsuperscripted \mathbf{f} is the flux interior to the element.

Appendix C

Runge-Kutta coefficients

C.1 Low-storage RK4

Following are the coefficients for a fourth-order low-storage RK method due to *Carpenter and Kennedy* [1994]. The algorithm is given in Chapter 3, Equation 3.48.

$$\begin{array}{lll}
 a_1 = 0 & b_1 = \frac{1432997174477}{9575080441755} & c_1 = 0 \\
 a_2 = -\frac{567301805773}{1357537059087} & b_2 = \frac{5161836677717}{13612068292357} & c_2 = \frac{1432997174477}{9575080441755} \\
 a_3 = -\frac{2404267990393}{2016746695238} & b_3 = \frac{1720146321549}{2090206949498} & c_3 = \frac{2526269341429}{6820363962896} \\
 a_4 = -\frac{3550918686646}{2091501179385} & b_4 = \frac{3134564353537}{4481467310338} & c_4 = \frac{2006345519317}{3224310063776} \\
 a_5 = -\frac{1275806237668}{842570457699} & b_5 = \frac{2277821191437}{14882151754819} & c_5 = \frac{2802321613138}{2924317926251}
 \end{array}$$

C.2 IMEX RK4

Following are the coefficients for an implicit-explicit RK method described in *Kennedy and Carpenter* [2003]. Specifically, we use their ARK4(3)6L[2]SA method, a 6-stage, fourth order IMEX RK method with a singly diagonally implicit stage. The algorithm is given in Chapter 3, Equation 3.51.

The coefficients are given in matrix form, so that a_{ij} is the $(i, j)^{\text{th}}$ entry of the matrix and b_j is the $(j)^{\text{th}}$ entry of \mathbf{b} .

C.2.1 Explicit stage coefficients

$$\mathbf{a}^{[\text{ex}]} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{13861}{62500} & \frac{6889}{62500} & 0 & 0 & 0 & 0 \\ -\frac{116923316275}{2393684061468} & -\frac{2731218467317}{15368042101831} & \frac{9408046702089}{11113171139209} & 0 & 0 & 0 \\ -\frac{451086348788}{2902428689909} & -\frac{2682348792572}{7519795681897} & \frac{12662868775082}{11960479115383} & \frac{3355817975965}{11060851509271} & 0 & 0 \\ \frac{647845179188}{3216320057751} & \frac{73281519250}{8382639484533} & \frac{552539513391}{3454668386233} & \frac{3354512671639}{8306763924573} & \frac{4040}{17871} & 0 \end{bmatrix}$$

C.2.2 Implicit stage coefficients

$$\mathbf{a}^{[\text{im}]} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{8611}{62500} & -\frac{1743}{31250} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{5012029}{34652500} & -\frac{654441}{2922500} & \frac{174375}{388108} & \frac{1}{4} & 0 & 0 \\ \frac{15267082809}{155376265600} & -\frac{71443401}{120774400} & \frac{730878875}{902184768} & \frac{2285395}{8070912} & \frac{1}{4} & 0 \\ \frac{82889}{524892} & 0 & \frac{15625}{83664} & \frac{69875}{102672} & -\frac{2260}{8211} & \frac{1}{4} \end{bmatrix}$$

C.2.3 Common coefficients

$$\mathbf{b} = \begin{bmatrix} \frac{82889}{524892} & 0 & \frac{15625}{83664} & \frac{69875}{102672} & -\frac{2260}{8211} & \frac{1}{4} \end{bmatrix}$$

$$\mathbf{c} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{83}{250} & \frac{31}{50} & \frac{17}{20} & 1 \end{bmatrix}$$

Bibliography

- Amestoy, P. R., I. S. Duff, J. Koster, and J.-Y. L'Excellent (2001), A fully asynchronous multifrontal solver using distributed dynamic scheduling, *SIAM Journal on Matrix Analysis and Applications*, 23(1), 15–41.
- Amestoy, P. R., A. Guermouche, J.-Y. L'Excellent, and S. Pralet (2006), Hybrid scheduling for the parallel solution of linear systems, *Parallel Computing*, 32(2), 136–156.
- Antonova, E. E. (2004), Magnetostatic equilibrium and current systems in the Earth's magnetosphere, *Advances in Space Research*, 33(5), 752–760.
- Antsaklis, P. J. (1997), *Linear systems*, McGraw-Hill.
- Appleton, E. V. (1928), The influence of the Earth's magnetic field on wireless transmission, *Proceedings of the Union Radio-Scientifique Internationale*, 1, 2–3.
- Armstrong, W. C. (1983), Recent advances from studies of the Trimpi effect, *Antarctic Journal United States*, 18, 281–283.
- Bailey, D. K. (1959), Abnormal ionization in the lower ionosphere associated with cosmic-ray flux enhancements, *Proceedings of the IRE*, 47(2), 255–266.
- Balay, S., W. D. Gropp, L. C. McInnes, and B. F. Smith (1997), Efficient management of parallelism in object oriented numerical software libraries, in *Modern Software Tools in Scientific Computing*, edited by E. Arge, A. M. Bruaset, and H. P. Langtangen, pp. 163–202, Birkhäuser Press.

- Balay, S., J. Brown, K. Buschelman, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang (2011), PETSc Web page, <http://www.mcs.anl.gov/petsc>.
- Balay, S., et al. (2010), PETSc users manual, *Tech. Rep. ANL-95/11 - Revision 3.1*, Argonne National Laboratory.
- Baumjohann, W., M. Blanc, A. Fedorov, and K.-H. Glassmeier (2010), Current systems in planetary magnetospheres and ionospheres, *Space Science Reviews*, 152, 99–134.
- Bécache, E., S. Fauqueux, and P. Joly (2003), Stability of perfectly matched layers, group velocities and anisotropic waves, *Journal of Computational Physics*, 188, no. 2, 399–433.
- Bell, T. F. (1984), The nonlinear gyroresonance interaction between energetic electrons and coherent VLF waves propagating at an arbitrary angle with respect to the Earth’s magnetic field, *Journal of Geophysical Research*, 89, 905–918.
- Bell, T. F. (1986), The wave magnetic field amplitude threshold for nonlinear trapping of energetic gyroresonant and Landau resonant electrons by nonducted VLF waves in the magnetosphere, *Journal of Geophysical Research*, 91, 4365–4379.
- Bell, T. F., and U. S. Inan (1981), Transient nonlinear pitch angle scattering of energetic electrons by coherent VLF wave packets in the magnetosphere, *Journal of Geophysical Research*, 86, 9047–9063.
- Bell, T. F., U. S. Inan, M. Platino, J. S. Pickett, P. A. Kossey, and E. J. Kennedy (2004), CLUSTER observations of lower hybrid waves excited at high altitudes by electromagnetic whistler mode signals from the HAARP facility, *Geophys. Res. Lett*, 31(6), 6811.
- Bell, T. F., U. S. Inan, D. Pidtychiy, P. Kulkarni, and M. Parrot (2008), Effects of plasma density irregularities on the pitch angle scattering of radiation belt electrons

- by signals from ground based VLF transmitters, *Geophysical Research Letters*, *35*, 19,103.
- Béranger, J.-P. (1994), A perfectly matched layer for the absorption of electromagnetic waves, *Journal of Computational Physics*, *114*, 185–200.
- Bilitza, D., and B. W. Reinisch (2008), International Reference Ionosphere 2007: improvements and new parameters, *Advances in Space Research*, *42*, 599–609.
- Birdsall, C. K., and A. B. Langdon (1985), *Plasma physics via computer simulation*, McGraw-Hill.
- Bittencourt, J., and M. Chrysafidis (1994), On the IRI model predictions for the low-latitude ionosphere, *Journal of Atmospheric and Terrestrial Physics*, *56*(8), 995–1009.
- Bortnik, J., R. M. Thorne, and N. P. Meredith (2008), The unexpected origin of plasmaspheric hiss from discrete chorus emissions, *Nature*, *452*, 62–66.
- Bortnik, J., L. Chen, W. Li, R. M. Thorne, N. Meredith, and R. B. Horne (2011), Modeling the wave power distribution and characteristics of plasmaspheric hiss, *Journal of Geophysical Research*, *116*.
- Brunini, C., A. Meza, F. Azpilicueta, A. Diaz, and M. A. van Zele (2004), A new ionosphere monitoring technology based on GPS, *United Nations Programme on Space Applications*, *15*, 51–64.
- Burtis, W. J., and R. A. Helliwell (1976), Magnetospheric chorus - Occurrence patterns and normalized frequency, *Planetary and Space Science*, *24*, 1007–1024.
- Bush, V., and H. Hazen (1931), The differential analyzer: a new machine for solving differential equations, *Journal of the Franklin Institute*, *212*(4), 447–488.
- Carpenter, D. L., et al. (2002), Small-scale field-aligned plasmaspheric density structures inferred from the Radio Plasma Imager on IMAGE, *Journal of Geophysical Research (Space Physics)*, *107*, 1258.

- Carpenter, M. H., and C. A. Kennedy (1994), Fourth-order 2N-storage Runge-Kutta schemes, *Nasa TM*, 109112(TM 109112).
- Chapman, S., and V. C. A. Ferraro (1931), A new theory of magnetic storms, *Terrestrial Magnetism and Atmospheric Electricity (Journal of Geophysical Research)*, 36, 77–97.
- Cheng, Z., and S. A. Cummer (2005), Broadband VLF measurements of lightning-induced ionospheric perturbations, *Geophysical Research Letters*, 32, L08,804.
- Cheng, Z., S. A. Cummer, H.-T. Su, and R.-R. Hsu (2007a), Broadband very low frequency measurement of D region ionospheric perturbations caused by lightning electromagnetic pulses, *Journal of Geophysical Research (Space Physics)*, 112(A11), A06318.
- Cheng, Z., S. A. Cummer, H.-T. Su, and R.-R. Hsu (2007b), Broadband very low frequency measurement of D region ionospheric perturbations caused by lightning electromagnetic pulses, *Journal of Geophysical Research (Space Physics)*, 112(A11), A06,318.
- Chevalier, M. W., and U. S. Inan (2004), A PML using a convolutional curl operator and a numerical reflection coefficient for general linear media, *IEEE Transactions on Antennas and Propagation*, 52, 1647–1657.
- Chevalier, T. W., U. S. Inan, and T. F. Bell (2008), Terminal Impedance and Antenna Current Distribution of a VLF Electric Dipole in the Inner Magnetosphere, *IEEE Transactions on Antennas and Propagation*, 56, 2454–2468.
- Chevalier, T. W., U. S. Inan, and T. F. Bell (2010), Fluid simulation of the collisionless plasma sheath surrounding an electric dipole antenna in the inner magnetosphere, *Radio Science*, 45, RS1010.
- Chew, W. C., and W. H. Weedon (1994), A 3D perfectly matched medium from modified Maxwell’s equations with stretched coordinates, *Microwave and Optical Technology Letters*, 7, 599–604.

- Church, S. R., and R. M. Thorne (1983), On the origin of plasmaspheric hiss - Ray path integrated amplification, *Journal of Geophysical Research*, 88, 7941–7957.
- Coisson, P., S. M. Radicella, L. Ciraolo, R. Leitinger, and B. Nava (2008), Global validation of IRI TEC for high and medium solar activity conditions, *Advances in Space Research*, 42(4), 770–775.
- Courant, R., E. Isaacson, and M. Rees (1952), On the solution of nonlinear hyperbolic differential equations by finite differences, *Communications on Pure and Applied Mathematics*, 5(3), 243–255.
- Cummer, S. A. (1997), Lightning and ionospheric remote sensing using VLF/ELF radio atmospherics, Ph.D. thesis, Stanford University.
- Daughton, W., J. Scudder, and H. Karimabadi (2006), Fully kinetic simulations of undriven magnetic reconnection with open boundary conditions, *Physics of Plasmas*, 13(7), 072,101.
- Davies, K. (1990), *Ionospheric radio*, Inst. of Electrical Engineers, London.
- De Schutter, B. (2000), Minimal state-space realization in linear system theory: an overview, *Journal of Computational and Applied Mathematics*, 121, 2000.
- De Schutter, B., and B. De Moor (1995), Minimal state space realization of MIMO systems in the max algebra, in *Proceedings of the 3rd European Control Conference*, pp. 411–416.
- Dettmering, D., R. Heinkelmann, and M. Schmidt (2011), Systematic differences between VTEC obtained by different space-geodetic techniques during CONT08, *Journal of Geodesy*, 85, 443–451.
- Dipeso, G., E. C. Morse, and R. W. Ziolkowski (1991), δf and particle simulations of parametric instabilities, *Journal of Computational Physics*, 96(2), 325–338.
- Edelsbrunner, H., and E. Mücke (1990), Simulation of simplicity: a technique to cope with degenerate cases in geometric algorithms, *ACM Transactions on Graphics*, 9, 66–104.

- Enge, P., T. Walter, S. Pullen, C. Kee, Y. C. Chao, and Y. J. Tsai (1996), Wide area augmentation of the Global Positioning System, *Proceedings of the IEEE*, 84(8), 1063–1088.
- Erturk, O., O. Arikan, and F. Arikan (2009), Tomographic reconstruction of the ionospheric electron density as a function of space and time, *Advances in Space Research*, 43(11), 1702–1710.
- Escudero, A., A. C. Schlesier, A. Rius, A. Flores, F. Rubek, G. B. Larsen, S. Syndergaard, and P. Hoeg (2001), Ionospheric tomography using Orsted GPS measurements - preliminary results, *Physics and Chemistry of the Earth A*, 26, 173–176.
- Foust, F. R., T. F. Bell, M. Spasojevic, and U. S. Inan (2011a), Discontinuous Galerkin particle-in-cell simulation of longitudinal plasma wave damping and comparison to the Landau approximation and the exact solution of the dispersion relation, *Physics of Plasmas*, 18, 2111.
- Foust, F. R., M. Spasojevic, T. F. Bell, and U. S. Inan (2011b), Modeling scattering from lightning-induced ionospheric disturbances with the discontinuous Galerkin method, *Journal of Geophysical Research*, 116.
- Galerkin, B. G. (1915), Series solution of some problems of elastic equilibrium of rods and plates, *Engineering News (Vestnik Inzhinierov)*, 19, 897–908.
- Gedney, S. D. (1996), An anisotropic perfectly matched layer-absorbing medium for the truncation of FDTD lattices, *IEEE Transactions on Antennas and Propagation*, 44, 1630–1639.
- Gedney, S. D., and B. Zhao (2010), An auxiliary differential equation formulation for the complex-frequency shifted PML, *IEEE Transactions on Antennas and Propagation*, 58, 838–847.
- Geuzaine, C., and J.-F. Remacle (2009), Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities, *International Journal for Numerical Methods in Engineering*, 79(11), 1309–1331.

- Gibson, W. C. (2008), *The method of moments in electromagnetics*, Chapman & Hall/CRC.
- Gilbert, E. G. (1963), Controllability and observability in multivariable control systems, *SIAM Journal on Control*, 2(1), 128–151.
- Graf, K. L., U. S. Inan, and M. Spasojevic (2011), Transmitter-induced modulation of subionospheric VLF signals: ionospheric heating rather than electron precipitation, *Journal of Geophysical Research*, 116.
- Gurnett, D. A., and A. Bhattacharjee (2005), *Introduction to plasma physics: with space and laboratory applications*, Cambridge University Press.
- Haque, N., U. S. Inan, T. F. Bell, J. S. Pickett, J.-G. Trotignon, and G. Facsko (2011), Cluster Observations of Whistler Mode Ducts and Banded Chorus, *Geophysical Research Letters*.
- Harrington, R. F. (1961), *Time-harmonic electromagnetic fields*, IEEE Press series on electromagnetic wave theory, IEEE Press.
- Hartley, D. (1999), EDSAC 1 and after - a compilation of personal reminiscences, <http://www.cl.cam.ac.uk/conference/EDSAC99/reminiscences/>.
- Hartree, D. R. (1931), The Propagation of Electromagnetic Waves in a Refracting Medium in a Magnetic Field, *Mathematical Proceedings of the Cambridge Philosophical Society*, 27(01), 143–162.
- Haselgrove, J. (1955), Ray theory and a new method for ray tracing, in *Physics of the Ionosphere*, p. 355.
- Hayakawa, M., and S. S. Sazhin (1992), Mid-latitude and plasmaspheric hiss - A review, *Planetary and Space Science*, 40, 1325–1338.
- Henry, F. S. (1980), Improved ray description of wave equations, *Physical Review Letters*, 45, 1897–1900.

- Hesthaven, J. S., and T. Warburton (2002), Nodal high-order methods on unstructured grids I. Time-domain solution of Maxwell's equations, *Journal of Computational Physics*, 181, 186–221.
- Hesthaven, J. S., and T. Warburton (2007), *Nodal discontinuous Galerkin methods: algorithms, analysis, and applications*, 1st ed., Springer Publishing Company, Incorporated.
- Ho, B. L., and R. E. Kalman (1966), Effective construction of linear state-variable models from input-output functions, *Regelungstechnik*, 14(12), 545–592.
- Hobbs, P. V. (2000), *Introduction to atmospheric chemistry: a companion text to basic physical chemistry for the atmospheric sciences*, Cambridge University Press.
- Hrennikoff, A. (1941), Solution of problems of elasticity by the framework method, *Journal of Applied Mechanics*, A8, 169–175.
- Huang, K. (2008), *Statistical mechanics*, Wiley India Pvt. Ltd.
- Hughes, T. J. R. (2000), *The finite element method: linear static and dynamic finite element analysis*, Dover Civil and Mechanical Engineering Series, Dover Publications.
- Hysom, D., and A. Pothen (1999), Efficient parallel computation of ILU(k) preconditioners, in *Proceedings of the 1999 ACM/IEEE conference on Supercomputing (CDROM)*, Supercomputing '99, ACM, New York, NY, USA.
- Hysom, D., and A. Pothen (2000), A scalable parallel algorithm for incomplete factor preconditioning, *SIAM Journal on Scientific Computing*, 22, 2194–2215.
- Inan, U. S., and M. Gołkowski (2010), *Principles of Plasma Physics for Engineers and Scientists*, Principles of Plasma Physics for Engineers and Scientists, Cambridge University Press.
- Inan, U. S., and R. A. Marshall (2011), *Numerical Electromagnetics: The FDTD Method*, Cambridge University Press.

- Inan, U. S., T. F. Bell, and J. V. Rodriguez (1991), Heating and ionization of the lower ionosphere by lightning, *Geophysical Research Letters*, *18*, 705–708.
- Inan, U. S., V. P. Pasko, and T. F. Bell (1996), Sustained heating of the ionosphere above thunderstorms as evidenced in “early/fast” VLF events, *Geophysical Research Letters*, *23*, 1067–1070.
- Inan, U. S., S. A. Cummer, and R. A. Marshall (2010), A survey of ELF and VLF research on lightning-ionosphere interactions and causative discharges, *Journal of Geophysical Research*, *115*.
- Jacobs, G. B., and J. S. Hesthaven (2006), High-order nodal discontinuous Galerkin particle-in-cell method on unstructured grids, *Journal of Computational Physics*, *214*, 96–121.
- Jacobs, G. B., and J. S. Hesthaven (2009), Implicit-explicit time integration of a high-order particle-in-cell method with hyperbolic divergence cleaning, *Computer Physics Communications*, *180*, 1760–1767.
- Jakowski, N., E. Sardon, and S. Schlueter (1998), GPS-based TEC observations in comparison with IRI95 and the European TEC model NTCM2, *Advances in Space Research*, *22*(6), 803–806.
- James, H. G., and T. F. Bell (1987), Spin modulation of spectrally broadened VLF signals, *Journal of Geophysical Research*, *92*, 7560–7568.
- Jin, J. M. (2002), *The finite element method in electromagnetics*, A Wiley-Interscience publication, Wiley.
- Johnson, M. P., and U. S. Inan (2000), Sferic Clusters Associated with Early/Fast VLF Events, *Geophysical Research Letters*, *27*, 1391.
- Johnson, M. P., U. S. Inan, S. J. Lev-Tov, and T. F. Bell (1999), Scattering pattern of lightning-induced ionospheric disturbances associated with early/fast VLF events, *Geophysical Research Letters*, *26*, 2363–2366.

- Jordanova, V. K., R. M. Thorne, W. Li, and Y. Miyoshi (2010), Excitation of whistler mode chorus from global ring current simulations, *Journal of Geophysical Research (Space Physics)*, *115*(A14), A00F10.
- Kanevsky, A., M. H. Carpenter, D. Gottlieb, and J. S. Hesthaven (2007), Application of implicit explicit high order Runge Kutta methods to discontinuous-Galerkin schemes, *Journal of Computational Physics*, *225*, 1753–1781.
- Karypis, G., K. Schloegel, and V. Kumar (1997), ParMETIS 1.0: parallel graph partitioning and sparse matrix ordering library, *Tech. Rep. TR-97-060*, Department of Computer Science, (also available electronically via <http://www.cs.umn.edu/~metis>).
- Katoh, Y., and Y. Omura (2006), A study of generation mechanism of VLF triggered emission by self-consistent particle code, *Journal of Geophysical Research (Space Physics)*, *111*(A10), A12207.
- Katoh, Y., and Y. Omura (2011), Amplitude dependence of frequency sweep rates of whistler mode chorus emissions, *Journal of Geophysical Research (Space Physics)*, *116*(A15), 7201.
- Kelley, M. C. (2009), *The Earth's ionosphere: plasma physics and electrodynamics*, International geophysics series, Academic Press.
- Kempf, K. (1961), *Electronic computers within the Ordnance Corps*, U.S. Army Ordnance Corps.
- Kennedy, C. A., and M. H. Carpenter (2003), Additive Runge-Kutta schemes for convection-diffusion-reaction equations, *Applied Numerical Mathematics*, *44*, 139–181.
- Kennel, C. F., and H. E. Petschek (1966), Limit on Stably Trapped Particle Fluxes, *Journal of Geophysical Research*, *71*, 1.

- Klimo, O., S. Weber, V. T. Tikhonchuk, and J. Limpouch (2010), Particle-in-cell simulations of laser-plasma interaction for the shock ignition scenario, *Plasma Physics and Controlled Fusion*, 52(5), 055,013.
- Koons, H. C. (1989), Observations of large-amplitude, whistler mode wave ducts in the outer plasmasphere, *Journal of Geophysical Research*, 94, 15,393–15,397.
- Kunitsyn, V. E., and E. D. Tereshchenko (2010), *Ionospheric tomography*, Physics of Earth and Space Environments, Springer.
- Landau, L. D. (1946), On the vibration of the electronic plasma, *Journal of Physics USSR*, 10, 25.
- Lassen, H. (1927), Über den Einflußdes Erdmagnetfeldes auf die Fortpflanzung der elektrischen Wellen der drahtlosen telegraphic in der Atmosphäre, *Elektrische Nachrichten-technik*, 4, 324–334.
- Lax, P., and B. Wendroff (1960), Systems of conservation laws, *Communications on Pure and Applied Mathematics*, 13(2), 217–237.
- Lax, P. D. (1958), Differential equations, difference equations and matrix theory, *Communications on Pure and Applied Mathematics*, 11(2), 175–194.
- Lax, P. D., and J. B. Keller (1951), The initial and maxed initial and boundary value problem for hyperbolic systems, *Technical report LAMS-1205*, Los Alamos National Laboratory, Los Alamos, NM, USA.
- Lax, P. D., and R. D. Richtmyer (1956, republished 2005), Survey of the stability of linear finite difference equations, in *Selected Papers Volume I*, edited by P. Sarnak and A. Majda, pp. 125–151, Springer New York.
- Lax, P. D., and B. Wendroff (1964, republished 2005), Difference schemes for hyperbolic equations with high order of accuracy, in *Selected Papers Volume I*, edited by P. Sarnak and A. Majda, pp. 284–301, Springer New York.

- Lehtinen, N. G., and U. S. Inan (2008), Radiation of ELF/VLF waves by harmonically varying currents into a stratified ionosphere with application to radiation by a modulated electrojet, *Journal of Geophysical Research (Space Physics)*, *113*(A12), 6301.
- Lehtinen, N. G., and U. S. Inan (2009), Full-wave modeling of transionospheric propagation of VLF waves, *Geophysical Research Letters*, *36*, 3104.
- Lehtinen, N. G., R. A. Marshall, and U. S. Inan (2010), Full-wave modeling of early VLF perturbations caused by lightning electromagnetic pulses, *Journal of Geophysical Research (Space Physics)*, *115*(A14), 0.
- LeVeque, R. J. (2002), *Finite volume methods for hyperbolic problems*, Cambridge texts in applied mathematics, Cambridge University Press.
- Lifshitz, E. M., L. D. Landau, and L. P. Pitaevskii (1994), *Statistical physics*, Course of theoretical physics, Pergamon Press.
- Lühr, H., and C. Xiong (2010), IRI-2007 model overestimates electron density during the 23/24 solar minimum, *Geophysical Research Letters*, *37*, 23,101.
- Madsen, N. K., and R. W. Ziolkowski (1990), A three-dimensional modified finite volume technique for Maxwell's equations, *Electromagnetics*, *10*(1-2), 147–161.
- Marshall, R. A., and U. S. Inan (2010), Two-dimensional frequency domain modeling of lightning EMP-induced perturbations to VLF transmitter signals, *Journal of Geophysical Research (Space Physics)*, *115*(A14), A00E29.
- Marshall, R. A., U. S. Inan, and W. A. Lyons (2006), On the association of early/fast very low frequency perturbations with sprites and rare examples of VLF backscatter, *Journal of Geophysical Research (Atmospheres)*, *111*(D10), D19108.
- Marshall, R. A., U. S. Inan, and T. W. Chevalier (2008a), Early VLF perturbations caused by lightning EMP-driven dissociative attachment, *Geophysical Research Letters*, *35*, 21,807.

- Marshall, R. A., U. S. Inan, and T. W. Chevalier (2008b), Early VLF perturbations caused by lightning EMP-driven dissociative attachment, *Geophysical Research Letters*, *35*, L21807.
- Marshall, R. A., U. S. Inan, and V. S. Glukhov (2010), Elves and associated electron density changes due to cloud-to-ground and in-cloud lightning discharges, *Journal of Geophysical Research (Space Physics)*, *115*(A14).
- McPherron, R. L. (1991), Physical processes producing magnetospheric substorms and magnetic storms, in *Geomagnetism*, edited by J. A. Jacobs, pp. 593–739.
- Mehrenberger, M., E. Violar, O. Hoenen, M. Campos Pinto, and E. Sonnendrücker (2006), A parallel adaptive Vlasov solver based on hierarchical finite element interpolation, *Nuclear Instruments and Methods in Physics Research A*, *558*, 188–191.
- Mitchell, C. N., L. Kersley, J. A. T. Heaton, and S. E. Pryse (1997), Determination of the vertical electron-density profile in ionospheric tomography: experimental results, *Annales Geophysicae*, *15*, 747–752.
- Moore, R. C., C. P. Barrington-Leigh, U. S. Inan, and T. F. Bell (2003), Early/fast VLF events produced by electron density changes associated with sprite halos, *Journal of Geophysical Research (Space Physics)*, *108*, 1363.
- Munz, C.-D., P. Omnes, R. Schneider, E. Sonnendrücker, and U. Voß (2000), Divergence correction techniques for Maxwell solvers based on a hyperbolic model, *Journal of Computational Physics*, *161*, 484–511.
- Nagano, I., K. Miyamura, S. Yagitani, I. Kimura, T. Okada, K. Hashimoto, and A. Y. Wong (1994), Full wave calculation method of VLF wave radiated from a dipole antenna in the ionosphere-analysis of joint experiment by HIPAS and Akebono satellite, *Electronics and Communications in Japan (Part I: Communications)*, *77*(11), 59–71.
- Northrop, T. G. (1963), Adiabatic Charged-Particle Motion, *Reviews of Geophysics and Space Physics*, *1*, 283–304.

- Nunn, D. (1974), A self-consistent theory of triggered VLF emissions, *Planetary and Space Science*, 22(3), 349 – 378.
- Nunn, D. (1990), The numerical simulation of VLF nonlinear wave-particle interactions in collision-free plasmas using the Vlasov hybrid simulation technique, *Computer Physics Communications*, 60(1), 1–25.
- Nygren, T. (1981), A simple method for obtaining reflection and transmission coefficients and fields for an electromagnetic wave in a horizontally stratified ionosphere, *Planetary and Space Science*, 29, 521–528.
- Olson, W. P. (1984), Introduction to the topology of magnetospheric current systems, in *Magnetospheric Currents*, edited by T. A. Potemra, pp. 49–62.
- Omidi, N., X. Blanco-Cano, C. T. Russell, H. Karimabadi, and M. Acuna (2002), Hybrid simulations of solar wind interaction with magnetized asteroids: General characteristics, *Journal of Geophysical Research (Space Physics)*, 107, 1487.
- Omura, Y., D. Nunn, H. Matsumoto, and M. J. Rycroft (1991), A review of observational, theoretical and numerical studies of VLF triggered emissions, *Journal of Atmospheric and Terrestrial Physics*, 53(5), 351–368.
- Omura, Y., Y. Katoh, and D. Summers (2008), Theory and simulation of the generation of whistler-mode chorus, *Journal of Geophysical Research (Space Physics)*, 113(A12), A04223.
- Pakula, W. A., P. F. Fougere, J. A. Klobuchar, H. J. Kuenzler, M. J. Buonsanto, J. M. Roth, J. C. Foster, and R. E. Sheehan (1995), Tomographic reconstruction of the ionosphere over North America with comparisons to ground-based radar, *Radio Science*, 30, 89–104.
- Pappert, R. A., and J. A. Ferguson (1986), VLF/LF mode conversion model calculations for air to air transmissions in the Earth-ionosphere waveguide, *Radio Science*, 21, 551–558.

- Pappert, R. A., and F. P. Snyder (1972), Some results of a mode-conversion program for VLF, *Radio Science*, 7, 913.
- Pasko, V. P., U. S. Inan, Y. N. Taranenko, and T. F. Bell (1995), Heating, ionization and upward discharges in the mesosphere due to intense quasi-electrostatic thundercloud fields, *Geophysical Research Letters*, 22, 365–368.
- Pasko, V. P., U. S. Inan, and T. F. Bell (1998), Ionospheric effects due to electrostatic thundercloud fields, *Journal of Atmospheric and Solar-Terrestrial Physics*, 60, 863–870.
- Poevlein, H. (1962), Sommerfeld-Runge law in three and four dimensions, *Physical Review*, 128, 956–964.
- Potemra, T. A. (1979), Current systems in the Earth’s magnetosphere - A review of U.S. progress for the 1975-1978 IUGG quadrennial report, *Reviews of Geophysics and Space Physics*, 17, 640–656.
- Poulsen, W. L., T. F. Bell, and U. S. Inan (1990), Three-dimensional modeling of subionospheric VLF propagation in the presence of localized D region perturbations associated with lightning, *Journal of Geophysical Research*, 95, 2355–2366.
- Poulsen, W. L., U. S. Inan, and T. F. Bell (1993), A multiple-mode three-dimensional model of VLF propagation in the earth-ionosphere waveguide in the presence of localized D region disturbances, *Journal of Geophysical Research*, 98, 1705–1717.
- Raymond, T. D., S. J. Franke, and K. C. Yeh (1994), Ionospheric tomography: its limitations and reconstruction methods, *Journal of Atmospheric and Terrestrial Physics*, 56(5), 637–657.
- Roden, J. A., and S. D. Gedney (2000), Convolution PML (CPML): An efficient FDTD implementation of the CFS-PML for arbitrary media, *Microwave and Optical Technology Letters*, 27(5), 334–339.
- Rodger, C. J., and R. J. McCormick (2006), Remote sensing of the upper atmosphere by VLF, in *Sprites, Elves and Intense Lightning Discharges*, *NATO Science Series*,

- vol. 225, edited by M. Füllekrug, E. A. Mareev, and M. J. Rycroft, pp. 167–190, Springer Netherlands.
- Roman, S. (2005), *Advanced Linear Algebra*, Graduate Texts in Mathematics, Springer.
- Ruffini, G., A. Flores, and A. Rius (1998), GPS tomography of the ionospheric electron content with a correlation functional, *IEEE Transactions on Geoscience and Remote Sensing*, 36(1), 143–153.
- Runge, C. (1912), *Graphical methods*, Columbia University lectures, Columbia University Press.
- Sadiku, M. N. O. (2000), *Numerical techniques in electromagnetics*, CRC Press.
- Santolík, O. (2008), New results of investigations of whistler-mode chorus emissions, *Nonlinear Processes in Geophysics*, 15, 621–630.
- Sarri, G., G. C. Murphy, M. E. Dieckmann, A. Bret, K. Quinn, I. Kourakis, M. Borghesi, L. O. C. Drury, and A. Ynnerman (2011), Two-dimensional particle-in-cell simulation of the expansion of a plasma into a rarefied medium, *New Journal of Physics*, 13(7), 073,023.
- Sazhin, S. S., and M. Hayakawa (1992), Magnetospheric chorus emissions - A review, *Planetary and Space Science*, 40, 681–697.
- Sechrist, C. F., Jr. (1974), Comparisons of techniques for measurement of D-region electron densities, *Radio Science*, 9, 137.
- Smith, R. L. (1961), Propagation Characteristics of Whistlers Trapped in Field-Aligned Columns of Enhanced Ionization, *Journal of Geophysical Research*, 66, 3699–3707.
- Soffer, A., and C. Stucchio (2008), A stable absorbing boundary layer for anisotropic waves, *ArXiv e-prints*.

- Solomon, J., N. Cornilleau-Wehrlin, A. Korth, and G. Kremser (1988), An experimental study of ELF/VLF hiss generation in the Earth's magnetosphere, *Journal of Geophysical Research*, *93*, 1839–1847.
- Sonnendrücker, E., F. Filbet, A. Friedman, E. Oudet, and J.-L. Vay (2004), Vlasov simulations of beams with a moving grid, *Computer Physics Communications*, *164*, 390–395.
- Sonnendrücker, E., M. Gutnic, M. Haefele, and J.-L. Lemaire (2005), Adaptive Vlasov simulations of intense beams, in *High intensity and high brightness hadron beams*, *American Institute of Physics Conference Series*, vol. 773, edited by R. W. Hasse, J. M. Lagniel, and I. Hofmann, pp. 155–157.
- Sonwalkar, V. S., and U. S. Inan (1989), Lightning as an embryonic source of VLF hiss, *Journal of Geophysical Research*, *94*, 6986–6994.
- Stockem, A., M. E. Dieckmann, and R. Schlickeiser (2010), PIC simulations of the temperature anisotropy-driven Weibel instability: analysing the perpendicular mode, *Plasma Physics and Controlled Fusion*, *52*(8), 085,009.
- Strang, G., and G. J. Fix (2008), *An analysis of the finite element method*, Wellesley-Cambridge Press.
- Strobel, D. F., T. R. Young, R. R. Meier, T. P. Coffey, and A. W. Ali (1974), The nighttime ionosphere: E region and lower F region, *Journal of Geophysical Research*, *79*, 3171–3178.
- Taflove, A., and S. C. Hagness (2005), *Computational electrodynamics: the finite-difference time-domain method*, 3rd ed., Artech House, Norwood.
- Tanaka, Y., M. Hayakawa, D. Lagoutte, F. Lefeuvre, and S. Tajima (1987), Spectral broadening of VLF transmitter signals and sideband structure observed on Aureol 3 satellite at middle latitudes, *Journal of Geophysical Research*, *92*, 7551–7559.

- Taranenko, Y. N., U. S. Inan, and T. F. Bell (1993), The interaction with the lower ionosphere of electromagnetic pulses from lightning: Excitation of optical emissions, *Geophysical Research Letters*, *20*, 2675–2678.
- Teixeira, F. L., and W. C. Chew (1998), General closed-form PML constitutive tensors to match arbitrary bianisotropic and dispersive linear media, *Microwave and Guided Wave Letters, IEEE*, *8*(6), 223–225.
- Tikhonov, A. N., and A. A. Samarskii (1960), Homogeneous difference schemes, *USSR Computational Mathematics and Mathematical Physics*, *1*(1), 5–67.
- Titova, E. E., V. I. di, V. E. Iurov, O. M. Raspopov, V. I. Trakhtengerts, F. Jiricek, and P. Triska (1984), Interaction between VLF waves and the turbulent ionosphere, *Geophysical Research Letters*, *11*, 323–326.
- Tsai, L.-C., C. H. Liu, W. H. Tsai, and C. T. Liu (2002), Tomographic imaging of the ionosphere using the GPS/MET and NNSS data, *Journal of Atmospheric and Solar-Terrestrial Physics*, *64*(18), 2003–2011.
- Uman, M. A., and D. K. McLain (1970), Lightning return stroke current from magnetic and radiation field measurements, *Journal of Geophysical Research*, *75*, 5143–5147.
- van der Vorst, H. A. (1992), BI-CGSTAB: a fast and smoothly converging variant of BI-CG for the solution of nonsymmetric linear systems, *SIAM Journal on Scientific and Statistical Computing*, *13*, 631–644.
- von Neumann, J. (1945), First draft of a report on the EDVAC (reprinted in 1993), *IEEE Annals of the History of Computing*, *15*, 27–75.
- Wait, J. R. (1996), *Electromagnetic Waves in Stratified Media*, Oxford University Press, USA.
- Weinberger, H. F. (1959), Error bounds in finite-difference approximation to solutions of symmetric hyperbolic systems, *Journal of the Society for Industrial and Applied Mathematics*, *7*(1), pp. 49–75.

- Whaley, R. C., and J. J. Dongarra (1997), Automatically tuned linear algebra software, *Tech. Rep. UT-CS-97-366*, University of Tennessee.
- Whaley, R. C., and J. J. Dongarra (1998), Automatically tuned linear algebra software, in *SuperComputing 1998: High Performance Networking and Computing*.
- Whaley, R. C., and J. J. Dongarra (1999), Automatically tuned linear algebra software, in *Ninth SIAM Conference on Parallel Processing for Scientific Computing*, cD-ROM Proceedings.
- Whaley, R. C., and A. Petitet (2005), Minimizing development and maintenance costs in supporting persistently optimized BLAS, *Software: Practice and Experience*, 35(2), 101–121.
- Whaley, R. C., A. Petitet, and J. J. Dongarra (2001), Automated empirical optimization of software and the ATLAS project, *Parallel Computing*, 27(1–2), 3–35.
- Yagitani, S., I. Nagano, K. Miyamura, and I. Kimura (1994), Full wave calculation of ELF/VLF propagation from a dipole source located in the lower ionosphere, *Radio Science*, 29, 39–54.
- Yee, K. (1966), Numerical solution of initial boundary value problems involving Maxwell’s equations in isotropic media, *IEEE Transactions on Antennas and Propagation*, 14, 302–307.
- Yeh, K. C., and C.-H. Liu (1972), *Theory of ionospheric waves*, International geophysics series, Academic Press.