# ABSTRACT

Title of Dissertation: ACCELERATION OF
KINETIC PLASMA SIMULATIONS
USING DATA-DRIVEN METHODS

Nathaniel Jared Barbour
Doctor of Philosophy, 2026

Dissertation Directed by: Professor James F. Drake
Department of Physics

Magnetic confinement fusion (MCF) is a promising candidate for a low-carbon electricity solution. MCF uses strong magnetic fields to confine hot, highly ionized plasmas for long enough times that nuclear fusion reactions can occur. Currently, the two most prominent MCF concepts are the tokamak and the stellarator, which confine plasmas in toroidal vessels. Due to strong temperature gradients, plasmas in both tokamaks and stellarators can exhibit kinetic small-scale turbulence, which drives unwanted macroscopic transport of heat and particles. The gyrokinetic simulations used to model this turbulence can be computationally intensive, so a valuable objective is to accelerate these simulations. A complementary goal is to ensure that simulation results are reliable for a given domain and resolution configuration.

In this thesis, we make two contributions in support of these initiatives. First, we explore the use of machine learning to reduce resolution requirements for kinetic plasma simulations, which would make them faster. We achieve this by developing a machine learning (ML) closure model for

velocity space in the one-dimensional Vlasov-Poisson system in a Fourier-Hermite spectral basis. The ML closure uses reservoir computing, a framework for training recurrent neural networks. We exploit the locality of interactions in the Hermite representation of velocity space to integrate large-scale dynamics with a kinetic solver, while using the ML closure model of the small-scale dynamics. In the linear limit, when the kinetic Fourier–Hermite solver is augmented with the reservoir closure, the closure permits a reduction of the velocity resolution, with a relative error within 2% for the Hermite moment where the reservoir closes the coupled hierarchy. In the strongly nonlinear regime of the Vlasov-Poisson system, the ML closure model more accurately resolves the low-order Fourier and Hermite spectra when compared with a naive closure by truncation and reduces the required velocity resolution by a factor of 16.

Second, we use correlation functions, heat flux and free energy spectra, and nonlinear energy transfer diagnostics to assess the convergence in heat flux for a set of electrostatic gyrokinetic turbulence simulations. We perform a parameter scan of temperature gradients, temperature ratios, domain extents, and resolutions in the plane perpendicular to the magnetic field lines simulated. Though no universal convergence threshold was found in correlation length or resolution, trends in energy and heat flux spectra point toward the potential to use spectra for the automatic identification of under-resolution during the execution of a simulation.

# ACCELERATION OF KINETIC PLASMA SIMULATIONS USING DATA-DRIVEN METHODS

by

Nathaniel Jared Barbour

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2026

Advisory Committee:
Professor James F. Drake, Chair
Dr. Ian G. Abel, Advisor
Dr. Matthew J. Landreman, Co-Advisor
Professor Thomas M. Antonsen, Jr.
Professor Alexander Philippov

© Copyright by
Nathaniel Jared Barbour
2026

# Dedication

In honor and remembrance of Professor William D. Dorland, Professor Nuno F. G. Louriero, George E. Barbour, Gloria J. Cross, and Mary H. Lewis.

# Acknowledgments

What a journey this has been. I take inspiration from Odysseus, who through wit, determination, and divine intervention, made his way home. Although each person encounters obstacles and difficulties on their journey, we overcome them through the support of our mentors, friends, and family. I am blessed to have received the support of many wonderful individuals. I apologize to anyone who I have inadvertently excluded.

First, I express my sincere gratitude for the patient mentorship and support of Prof. Bill Dorland, whose innovative expertise and generous spirit brought me to the University of Maryland. I will always remember the way he welcomed me into computational plasma physics with open arms, sharing a warm, lively discussion with me and my father in his office during a visit to UMD as an undergrad. Moments like that can shape the trajectory of a person's life. I am also grateful for the mentorship of Prof. Nuno Loureiro, whose vibrant enthusiasm inspired me to become a better ambassador for the field. I shall endeavor to honor their memory in my career.

I thank Dr. Ian Abel and Dr. Matt Landreman for their generous offers to support me, to aid me in writing a paper on my work with Prof. Dorland, and to advise me on a final project at UMD. I also thank Prof. Jim Drake for volunteering to bear the additional responsibility of serving as the chair of my dissertation committee. I appreciate the time and effort of Prof. Tom Antonsen, Jr., and Prof. Sasha Philippov in serving on the committee. I also sincerely thank the administrators of IREAP for their tireless effort on the behalf of the faculty and students in the unit.

I also thank Dr. Rahul Gaur, Byoungchan Jang, Anna Fitzmaurice, Zheng-Yang Tan, John Kappel, Siena Hurwitz, Landry Horimbere, Patrick Kim, Alex Wiedman, Prof. Alan Kaptanoglu, Dr. Stefan Buller, Dr. Rory Conlin, and Dr. Todd Elder for their supportive companionship and camaraderie during our time conducting research at UMD.

I want to thank Capt. Royce James, Sara Negussie, Dr. Arturo Dominguez, Dr. Nick Murphy, Prof. David Newman, Dr. Mel Abler, Dr. Paul Keiter, Caira Anderson, Prof. Steffi Diem, Jaela Whitfield, Dr. Marissa Adams, Dr. Eli Parke, Dr. Luis Delgado-Aparicio, Dr. Ahmed Diallo, and Amelia Chambliss for reinforcing the value of solidarity and community by coming to my aid in the wake of the loss I experienced over the years.

I wish to thank Prof. Greg Hammett, Dr. Noah Mandell, Dr. Jimmy Juno, Dr. Mike Martin, Prof. Elizabeth Paul, Prof. Rogerio Jorge, Prof. Paulo Alves, and Prof. Frederico Fiúza for their encouragement and insightful advice in pursuit of a career in computational plasma physics for magnetic fusion. I am passionate about science policy, and I am most grateful to Prof. Jim Gates, Prof. Rosina Bierbaum, and Prof. Frank von Hippel for giving me the foundational training and confidence necessary to engage in policy advocacy and outreach initiatives.

I thank my undergraduate advisors and mentors, Prof. Charles Baltay, Prof. Egemen Kolemen, Dr. Elena Belova, Prof. Stephen Pitti, Prof. Alicia Schmidt Camacho, Prof. John Harris, Prof. Witold Skiba, Prof. Charles Brown, and Jonathan Hamill. Their guidance during the formative stages of my career was instrumental in preparing me for this moment.

I thank Dr. Leila Wahab, Abril Brugo, Jacques Baynes, Jr., Isaiah Baynes, Jensynia Baynes, Justin Raine, Christian Lee, Fariha Jawed, and Billy Graves for their supportive friendship and the cherished memories we have made together. They have helped me to find the silver linings in the clouds and the nourishment in the rain that inevitably comes.

Finally, I thank my wonderful family, with special thanks to my parents, Edward Barbour and Prof. Vanessa Browne-Barbour, my aunts Jacalyn Barbour and Victoria Browne, my uncle Tony Jones, my cousin Dontae Adyelott, and my grandparents George Barbour, Gloria Barbour, and Mary Looper. Their unceasing love and nurturing support for my dreams and their belief in

my abilities have granted me safe passage on this journey. They each instilled in me the value of education and inspired me to dedicate my efforts toward the betterment of humanity.

# Table of Contents


Dedication ii

Acknowledgements iii

Table of Contents vi

List of Tables viii

List of Figures ix

List of Abbreviations xiv

Chapter 1: Introduction 1
- 1.1 Motivation: Energy and Climate . . . 1
- 1.2 Nuclear Fusion . . . 3
- 1.3 Overview of the Thesis . . . 6

Chapter 2: Theoretical Background: Kinetic Theory 9
- 2.1 Defining our plasmas . . . 9
- 2.2 Electrostatic Gyrokinetic System with Adiabatic Electrons . . . 12
  - 2.2.1 Gyrokinetic orderings . . . 13
  - 2.2.2 The Gyroaverage . . . 17
  - 2.2.3 Zeroth order of the gyrokinetic equation . . . 19
  - 2.2.4 First order of the kinetic equation . . . 22
  - 2.2.5 Second order of the gyrokinetic equation . . . 25
  - 2.2.6 Quasineutrality: Determining Phi . . . 28
- 2.3 1-D Vlasov-Poisson System . . . 29

Chapter 3: Numerical Methods: The Eulerian Pseudo-spectral Approach 34
- 3.1 Electrostatic Gyrokinetics . . . 41
- 3.2 Vlasov-Poisson System . . . 60
  - 3.2.1 Hermite Moment Formalism . . . 60
  - 3.2.2 Vlasov Equation in the Hermite basis . . . 61
  - 3.2.3 Pseudo-spectral in Fourier Space . . . 63
  - 3.2.4 Poisson's Equation in the Fourier-Hermite basis . . . 65
- 3.3 Summary . . . 66

Chapter 4: Machine-learning Closure of Vlasov-Poisson Dynamics in Fourier-Hermite Space 67
- 4.1 Motivation . . . 67
  - 4.1.1 The Closure Problem . . . 71
- 4.2 Machine-Learning Closure: Reservoir Computing . . . 72

4.2.1 Closure Integration . . . 75
4.2.2 Computational Complexity . . . 76
4.3 Closure Results . . . 79
4.3.1 Linearized Vlasov-Poisson . . . 79
4.3.2 Weakly Nonlinear Vlasov-Poisson . . . 83
4.3.3 Strongly Nonlinear Vlasov-Poisson . . . 87
4.4 Summary and Conclusion . . . 91

Chapter 5: Progress Toward Automatic Identification of Under-resolved Gyrokinetic Turbulence Simulations: Data-Driven Heuristics 94
5.1 Simulation details . . . 96
5.2 Correlation Functions . . . 99
5.2.1 Parallel Correlations: Critical Balance . . . 99
5.2.2 Perpendicular Correlations . . . 102
5.3 Resolution . . . 118
5.3.1 Analysis of Heat Flux and Free Energy Spectra . . . 119
5.3.2 Nonlinear Energy Transfer Diagnostic . . . 122

Chapter 6: Summary and Discussion 129

Appendix A: Appendix to ML Closure Study 134
A.1 Hermite Resolution Convergence . . . 134
A.2 Time Evolution of Low-Order Moments . . . 134

Bibliography 138

# List of Tables

A.1 Damping rates calculated from the initial-value simulations in Figures A.3 – A.6. Here, $\epsilon$ is the amplitude of the initial cosine density perturbation, and $m$ is the Hermite moment number. Damping rates were calculated by solving a linear regression problem for the natural logarithm of the local maxima in each time series. 136

# List of Figures


3.1 a) Cosine wave superimposed with fine-scale spatial oscillations evaluated at 512 equally spaced grid points on the domain $x \in [0, 2\pi]$: $f(x) = \cos(x) + 0.2\cos(16x) + 0.1\cos(64x)$. b) Amplitudes of the positive wavenumbers $k$ for the Fourier transform of the waveform in a). . . . . 38

3.2 a) Inverse Fourier transform of the Fourier Series in b), which truncates the Fourier series in Figure 3.1 b) by setting all Fourier amplitudes for $|k| > 32$ to zero. The truncation removes the finest scale oscillations that resulted from the $0.1\cos(64x)$ term. . . . . 39

4.1 Diagram of the reservoir computing closure model, integrated with the kinetic moment solver. . . . . 75

4.2 High-velocity-resolution ($M = 1025$) baseline time series of Hermite amplitudes for the driven, linearized Vlasov-Poisson system. Energy is injected as a density perturbation in the $m = 0$ Hermite moment and advects to higher moments through linear Landau damping. No hypercollisional regularization is applied to this example, and hundreds of Hermite moments are required to prevent numerical reflection at the high-$m$ boundary from impacting the low-$m$ spectrum. . . . . 79

4.3 Time-averaged Hermite spectra for the ML closure model compared to the high-velocity-resolution ($M = 1025$) baseline, closure-by-truncation with three different coefficients ($\nu_m$) for hypercollisional regularization, and the theoretical $m^{-1/2}$ scaling. The ML closure model shows strong agreement with the baseline simulation, while no value of $\nu_m$ produces an accurate spectrum. . . . . 79

4.4 Comparison between baseline numerical solution, ML closure, and theoretical damping rate of the Fourier-Hermite amplitude for an initial cosine density perturbation. The low-amplitude perturbation shows strong agreement with the theoretical damping rate. When augmented with the ML closure, the moment solver continues to capture the behavior well at a lower Hermite resolution of $M = 4$, as opposed to the $M = 17$ baseline. . . . . 85

4.5 Hermite spectra of the high-velocity-resolution ($M = 17$), truncated ($M = 4$) simulations, and ML closure for the initial value problem in Figure 4.4 . The spectra are averaged over time and Fourier wavenumber. The ML closure model permits a low-resolution simulation to accurately resolve the Hermite spectrum. . . . . 85

4.6 Fourier spectra of the simulations in Figure 4.5 . In this weakly nonlinear regime, the majority of the energy remains in the boxscale mode. Each reduced Hermite resolution model properly captures the rapid energy decay across $k$. . . . . 85

4.7 Simulated Fourier-Hermite amplitude for a high initial amplitude (18% of background) cosine density perturbation. The dynamics exhibit strongly nonlinear behavior, attenuating the damping of the mode. As in the low-amplitude case, the ML closure model captures the frequency and amplitude of the wave well. . . . . 89

4.8 Hermite spectra of the high-velocity-resolution ($M = 65$), truncated ($M = 4$) simulations, and ML closure for the initial value problem in Figure 4.7 averaged over $k$ and $t$. While both the ML closure model and truncated simulation agree with the high-resolution direct numerical simulation, the ML closure model shows closer agreement. . . . . 89
4.9 Fourier spectra of the simulations in Figure 4.8. The ML closure resolves the wavenumber spectrum more accurately than the truncated simulations, improving simulation accuracy at low resolution in velocity space. . . . . 90

5.1 Visualizations of the flux surfaces for the ITG simulations in this thesis in cylindrical $(R, Z)$ coordinates. The cyclone base case uses a circular flux surface (left) constructed using a local Miller parameterization [Miller et al., 1998]. For W7-X, we choose the $s = 0.64$ flux surface from a VMEC equilibrium (right). Poloidal cross sections are shown at four equally spaced locations in toroidal angle $\zeta$ for one field period. . . . . 98
5.2 Parallel correlation function for a GX simulation of the cyclone base case using parameters similar to those in [Barnes et al., 2019]. The dashed line corresponds to the critical balance scaling law $(\ell_\parallel/R) \sim (\ell_\perp/\rho_i)^{4/3}$, where we take $\overline{\Delta\theta}$ as a measure of $\ell_\parallel/R$ and $k_y\rho_i$ as a measure of $\ell_\perp/\rho_i$. . . . . 101
5.3 Parallel correlation function for an ITG turbulence simulation of the $\alpha = 0$ field line on the flux surface in the right of Figure 5.1. As in Figure 5.2, the dotted line corresponds to the crtical balance scaling. The stellarator case indicates that the turbulence shows weak correlation in $\theta$. . . . . 101
5.4 Electric potential in the cyclone base case, evaluated at the outboard midplane 1,000 timesteps (left) and 5,000 timesteps (right) after initialization. . . . . 103
5.5 Heat flux time trace for the cyclone base case. The initial transient ends by $t(v_{ti}/a) = 100$. . . . . 104
5.6 Electric potential from the left plot of Figure 5.4, evaluated as a function of $y$ at $x = 124.8$. . . . . 104
5.7 $x$-averaged $x$ correlation function for the potential in Figure 5.4. . . . . 105
5.8 $y$-averaged $x$ correlation function for the potential in Figure 5.4. . . . . 105
5.9 Correlation function in $x$ averaged over $x$ and $t$ for the cyclone base case. . . . . 106
5.10 Correlation function in $x$ averaged over $y$ and $t$ for the cyclone base case. . . . . 106
5.11 Correlation function in $y$ averaged over $x$ and $t$ for the cyclone base case. This is a clear indication of zonal-flow-dominated turbulence, as there is a strong correlation in $y$ at all displacements. . . . . 107
5.12 Correlation function in $y$ averaged over $y$ and $t$ for the cyclone base case. As in Figure 5.11, there is clear evidence of a zonal flow. . . . . 107
5.13 Correlation function in $x$ after removing the $ky = 0$ mode of the potential and averaging over $x$ and $t$ for the cyclone base case. A clear peak emerges around zero displacement. . . . . 108
5.14 Correlation function in $y$ after removing the $ky = 0$ mode of the potential and averaging over $x$ and $t$ for the cyclone base case. . . . . 108

5.15 Correlation function in $x$ of the potential, averaged over $x$, $y$, and $t$ for the cyclone base case. The $k_y = 0$ zonal mode has a finite wavelength in $x$ that results in negative correlations at large displacements. . . . . 109
5.16 Correlation function in $y$, averaged over $x$, $y$, and $t$ for the cyclone base case. The zonal mode leads to positive displacements at all displacements. . . . . 109
5.17 Correlation function in $x$ after removing the $ky = 0$ mode of the potential and averaging over $x$ and $t$ for the cyclone base case. As expected, the correlation of the fluctuations decays toward zero. . . . . 110
5.18 Correlation function in $y$ after removing the $ky = 0$ mode of the potential and averaging over $x$ and $t$ for the cyclone base case. Removing the zonal flow reveals oscillations in the correlation of the fluctuations, but these decay toward zero near the maximum displacements. . . . . 110
5.19 Correlation lengths for the cyclone base case as a function of box size in $y$ for normalized temperature gradient $a/L_T = 5$. The blue points correspond to correlations in $x$, while the orange points are for correlations in $y$. Round markers indicate integrals of the correlation function with respect to displacement. Star markers represent integrals of the absolute value of the correlation function, which penalizes the resulting correlation length for potential negative correlations. For the cyclone base case, $x$ correlation lengths are consistently longer than the $y$ correlation lengths after removing the zonal flows. . . . . 113
5.20 Log-log plot of heat fluxes for the cyclone base case for a parameter scan in $y$ box size and temperature gradient. Uncertainty in the heat flux is inversely proportional with box size, despite heat flux increasing with box size. . . . . 115
5.21 Log-linear plot of heat flux as a function of effective box size for the same parameter scan over temperature gradient as in Figure 5.20. For $a/L_T < 5$, the heat flux appears to converge when the effective box size is at least ten times longer than $\ell_{c,x}$, the correlation length in $x$. . . . . 115
5.22 Log-log plot of heat flux as a function of box size $L_y$ for the cyclone base case at five values of the ratio between the electron and ion temperatures $T_e/T_i$. Although the heat flux increases with the temperature ratio, the perpendicular domain size requirement appears to be independent of the temperature ratio. . . . . 116
5.23 Log-log plot of heat flux as a function of box size normalized by the correlation length in $x$. For all values of the temperature ratio, a perpendicular domain that measures roughly five correlation length is sufficient to calculate an accurate heat flux. . . . . 116
5.24 Log-log plot of heat flux as a function of box size for a parameter scan in temperature gradient $a/L_T$. When compared to the cyclone base case, the W7-X simulations converge to an accurate heat flux at a smaller box size. Interestingly, a domain that is too small leads to an overestimate of the heat flux. . . . . 117
5.25 Log-linear plot of heat flux compared to effective box size for a scan over $a/L_T$ for the W7-X flux tube. Somewhere between four and five correlation lengths in $x$ seem to yield the converged heat flux. . . . . 117

5.26 Heat flux time traces for five simulations of the cyclone base case at resolutions $Ny \in \{16, 32, 64, 128, 256\}$, all with temperature gradient $a/L_T = 2.5$. The linear instability saturates approximately 100 $t(v_t/a)$ later for $N_y = 16$ than in the high-resolution $N_y = 256$ simulation. . . . . 121
5.27 Resolution convergence study comparing heat flux to resolution in $y$ for three values of the temperature gradient $a/L_T$. For $a/L_T = 2.5$ and 3.5, the heat flux appears to converge by $N_y = 64$, while more resolution is necessary for $a/L_T = 5$. 121
5.28 Heat flux spectra as a function of $k_y$ for the five simulations of the cyclone base case in Figure 5.26, plotted on a log-log scale. As the low resolution of $N_y = 16$, an inertial range fails to form. The $N_y = 128$ and $N_y = 256$ spectra strongly agree with each other, and the $N_y = 64$ spectrum agrees at low $k_y$. . . . . 122
5.29 Log-log plot of the free energy spectra for the resolution scan shown in Figures 5.26 and 5.28. Unlike the heat flux spectra, the free energy spectrum saturates at high $k_y$ in the $N_y = 256$ case, which may indicate that additional dissipation may be necessary. . . . . 122
5.30 Perpendicular waveband shells for the cyclone base case evaluated at $\theta = 0$ (left) and $\theta = \pi/2$ (right). Magnetic shear couples different $(k_x, k_y)$ pairs together at different points along the field line. . . . . 124
5.31 Nonlinear energy transfer function $\mathcal{P}(K|P)$ for the $N_y = 16$ simulation of the cyclone base case with $a/L_T = 2.5$ and wavebands spaced by the dyadic progression $k_{n+1} = 2k_n$. . . . . 126
5.32 Nonlinear energy transfer function $\mathcal{P}(K|P)$ for the $N_y = 32$ simulation of the cyclone base case with $a/L_T = 2.5$. . . . . 126
5.33 Nonlinear energy transfer function $\mathcal{P}(K|P)$ for the $N_y = 64$ simulation of the cyclone base case with $a/L_T = 2.5$. . . . . 126
5.34 Nonlinear energy transfer function $\mathcal{P}(K|P)$ for the $N_y = 128$ simulation of the cyclone base case with $a/L_T = 2.5$. . . . . 126
5.35 Nonlinear energy transfer function $\mathcal{P}(K|P)$ for the $N_y = 256$ simulation of the cyclone base case with $a/L_T = 2.5$. . . . . 127

A.1 Convergence study of the Hermite spectra for the strongly nonlinear initial-value problem from Section 4.3.3 . We solve (3.94) - (3.97) without the ML closure at increasing levels of resolution in velocity space. We selected $M = 65$ moments for the high-resolution baseline case in Section 4.3.3 , as that case is the first to show convergence in the density and momentum moments. . . . . 135
A.2 Root-mean-square error in the low-order ($m \in [0, 1, 2, 3]$) Hermite spectra at the resolutions plotted in Figure A.1 , as compared to the $M = 513$ case. The $M = 17$ and $M = 33$ cases have similar errors, and exponential convergence occurs afterward. 135
A.3 Time trace of the $m = 1$ moment for the low-amplitude initial-value case in Section 4.3.2 . . . . . 137
A.4 Time trace of the $m = 2$ moment for the low-amplitude initial-value case in Section 4.3.2 . . . . . 137
A.5 Time trace of the $m = 1$ moment for the high-amplitude initial-value case in Section 4.3.3 . . . . . 137

A.6 Time trace of the $m = 2$ moment for the high-amplitude initial-value case in Section 4.3.3 . . . . . 137

# List of Abbreviations

| | |
|---|---|
| IREAP | Institute for Research in Electronics and Applied Physics |
| UMD | University of Maryland, College Park |
| MCF | Magnetic confinement fusion |
| MHD | Magnetohydrodynamic(s) |
| IVP(s) | Initial-value problem(s) |
| PDE(s) | Partial differential equation(s) |
| ML | Machine learning |
| AI | Artificial intelligence |
| GPU | Graphics processing unit |
| CUDA | Compute Unified Device Architecture |
| HIP | Heterogeneous-compute Interface for Portability |
| W7-X | Wendelstein 7-X |
| DIII-D | Doublet III-D |
| ITG | Ion temperature gradient |

# Chapter 1: Introduction

In this chapter, we provide a brief introduction to magnetic confinement fusion. We discuss the role of turbulence simulations in advancing fusion reactor design. We describe some of the features that make numerical simulations of magnetic fusion plasmas a challenging problem, and we give a small primer on turbulence modeling.

## 1.1 Motivation: Energy and Climate

From the dawn of civilization, energy has been pivotal to societal development. Even the ubiquitous wheel, a Neolithic invention, can be considered as an energy efficiency technology, as it reduces the energy required to transport materials across a distance. The invention of the steam engine catalyzed the Industrial Revolution. By burning fuel to heat water into pressurized steam that could spin turbines, these engines provided the efficient energy needed to power textile mills, mines, and manufacturing factories, eventually leading to global development. The combination of steam power and the dynamo made feasible the generation of electricity at scale via centralized power plants.

Though expanding access to electricity has undoubtedly shaped the modern world, it has also contributed to climate change. Global electricity production has grown from approximately ten terawatt-hours in 1900 to roughly fifteen thousand terawatt-hours in 2000 [Smil, 2017b]. Rising demand for electricity raises questions about the future of how to generate it. Currently, the majority of the energy used to produce electricity is derived from burning fossil fuels. This process produces carbon dioxide and other greenhouse gases, which accelerates climate change, potentially

impacting the habitability of affected areas over the course of the century and may drive migration [IPCC, 2023]. Limiting warming to well-below $2^{\circ}$ C above pre-industrial levels by the year 2050 requires $93 - 97$ percent of global electricity production to come from low-carbon sources [IPCC, 2023]. The status quo of relying on fossil fuels for electricity production is incompatible with the goal of avoiding systemic changes to the climate.

Current renewable, low carbon approaches to electricity production have drawbacks that limit their effectiveness. Geothermal power uses thermal energy stored beneath Earth's surface as the energy source for a thermal cycle that drives generator turbines. Wind and hydroelectric turbines reclaim mechanical energy from naturally occurring air and water currents. These technologies rely on the availability of geographic features and weather patterns and cannot serve as the foundation of the energy supply. In recent years, solar power has grown in its share of the energy mix. Solar power uses the photovoltaic effect to directly convert incoming electromagnetic radiation from the sun into an electric current [Smil, 2017a]. Although innovations in photovoltaic systems have led to consistent declines in costs over the past five decades [Kavlak et al., 2025], the intermittent availability of solar radiation impedes its ability to consistently meet electricity demand. Pairing solar systems with storage technologies, including batteries, pumped hydro, and molten salt storage, can potentially mitigate the intermittency problem, but these storage technologies each come with geographical or greenhouse gas emissions limitations [EsmaeiliShayan, 2025]. A transition to $100\%$ renewable energy for electricity production may be achievable, but it would require continued technological breakthroughs in efficiency, cost reductions, stability, and storage [Al-Shetwi et al., 2024].

## 1.2 Nuclear Fusion

Nuclear fusion is a promising solution to some of the challenges associated with renewables and nuclear fission. Fusion is the process of fusing lightweight atomic nuclei, often isotopes of hydrogen, to form heavier nuclei. This process releases significant amounts of energy. In fact, fusion is the most important source of energy that powers stars, including the sun [Bethe, 1939]. Most fusion reactor concepts focus on the deuterium-tritium fusion reaction

$$ {}^{2}_{1}\mathrm{H} + {}^{3}_{1}\mathrm{H} \rightarrow {}^{4}_{2}\mathrm{He}(3.5\,\mathrm{MeV}) + {}^{1}_{0}\mathrm{n}(14.1\,\mathrm{MeV}), \tag{1.1} $$

which fuses the hydrogen isotopes deuterium and tritium into an alpha particle and releases a high-energy neutron. An advantage of nuclear fusion is its energy density, with an advantage of a factor of four in comparison to fission and around a million in comparison to fossil fuels [Freidberg, 2007]. Fusion power plants have the potential to operate with high uptime and to provide firm, baseload power to the grid. Fusion likely will face fewer regulatory barriers to deployment at scale than fission. An example in the United States is the ADVANCE Act, folded into the Fire Grants and Safety Act, which streamlined fusion regulation by codifying the Nuclear Regulatory Commission's decision to regulate fusion facilities under a less restrictive byproduct materials framework than the more burdensome utilization facility framework it applies to fission power plants [118th United States Congress, 2024]. Fusion has no risk of criticality or meltdown unlike fission, and none of its byproducts are greenhouse gases.

Fusion reactions require significant amounts of energy to initiate. Positively charged ions must have a sufficient amount of energy to overcome the Coulomb force so that the strong nu-

clear force can fuse them together. Without access to the high gravitational forces of stars, fusion reactors instead must heat gases to extreme temperatures on the order of $10\,\mathrm{keV}$, which maximizes the cross section of the deuterium-tritium fusion reaction [Wurzel and Hsu, 2022]. At these temperatures, the gases ionize into plasmas that must be confined by methods other than material walls.

The two main branches of fusion confinement techniques are inertial confinement and magnetic confinement. Inertial confinement uses high-energy laser pulses ($\sim 2\,\mathrm{MJ}$) to form symmetric shock waves that compress and heat the plasma. Magnetic confinement uses strong magnetic fields ($\sim 2 - 12\,\mathrm{T}$) to constrain the charged ions and electrons to orbit around magnetic field lines. Both approaches have their merits. Inertial confinement is the only approach that has achieved "scientific breakeven" to date, meaning that in several experiments, fusion reactions released more energy than the laser energy required to heat and confine the plasma [Hurricane et al., 2024]. In contrast, magnetic confinement heats the plasma more efficiently than inertial confinement [Wurzel and Hsu, 2022], potentially signaling a faster path to commercialization once ignition has been achieved. For this reason, magnetic confinement fusion (MCF) is arguably the most promising approach.

In magnetic confinement fusion, the two main designs are the tokamak and stellarator. These machines are toroidally shaped chambers that use electromagnetic coils to create magnetic fields that form nested surfaces called "flux surfaces", where the magnetic field lines are tangent to the surfaces. The most simple toroidal magnetic field geometry would have magnetic field lines trace circles around a central column. However, magnetic drifts render this configuration to be unsuitable for confinement, as the charged particles would drift away from their original flux surfaces. Instead, the field lines must twist as they rotate around the torus. The pitch in the field lines is called the rotational transform. Tokamaks and stellarators differ in their approaches to generating this

rotational transform. Tokamaks maintain an axisymmetric shape so that their flux surfaces are symmetric in the toroidal angle. To achieve both axisymmetry and rotational transform, tokamaks rely on toroidal currents in the plasma that generate poloidal magnetic fields. Stellarators instead incorporate the twists into the magnetic fields generated by external electromagnetic coils. This allows them to forgo plasma currents and the family of current-driven instabilities that come with them.

Calculations of the power balance between energy inputs to heat the plasma and the energy yield of fusion reactions results in a condition called the Lawson criterion, which determines the density, temperature, and energy confinement time required to achieve self-sustaining fusion reactions [Lawson, 1957]. For deuterium-tritium MCF with equal ion and electron temperatures of $14\,\mathrm{keV}$, self-sustaining fusion, often called ignition, occurs at a fusion triple product of

$$n_i T \tau_E = 2.9 \times 10^{21}\,(\mathrm{m}^{-3} \cdot \mathrm{keV} \cdot \mathrm{s}), \tag{1.2}$$

where $n_i$, $T$, and $\tau_E$ are the ion density, temperature, and energy confinement time respectively [Wurzel and Hsu, 2022]. Although, no MCF experiment has achieved ignition, exponential improvements in the fusion triple product have been achieved over the last six decades. Multiple machines are currently under construction that are projected to achieve break-even energy yields by 2040. The peak of the deuterium-tritium cross section largely determines the ideal operating temperature for optimizing the triple product. Tokamaks and stellarators face empirical limits to their densities that constrain their operating spaces [Greenwald et al., 1988; Sudo et al., 1990]. These factors motivate efforts to increase the energy confinement time.

One of the core contributors to energy loss in MCF is the transport of heat driven by micro-

turbulence [Yoshida et al., 2025]. In principle, engineering experiments could iteratively optimize the magnetic geometry of tokamaks and stellarators to be robust to turbulence. However, the associated capital costs for construction and operation would be infeasible. Numerical simulation and modeling of turbulence is a vital tool that bridges the gap between analytical scaling approximations and empirical results. With densities on the order of $10^{20}\,\mathrm{m}^{-3}$, kinetic many-body first-principles simulations of the plasma dynamics would be computationally infeasible. Thankfully, there is a rigorous limit of the statistical kinetic description of plasmas that reduces computational requirements. This model is called gyrokinetics, and we derive the gyrokinetic system in Chapter 2. Gyrokinetic simulations of core turbulence in MCF plasmas have been validated against experiments [White, 2019; Yoshida et al., 2025].

## 1.3 Overview of the Thesis

Although gyrokinetic simulations are many orders of magnitude faster than many-body simulations that track each individual particle, we aspire to make them faster. Individual simulations that use further physics approximations, including the adiabatic electron response and static magnetic fields, can still require tens of minutes to complete on state-of-the-art hardware. With the shift toward multi-scale transport simulations and surrogate model development, workflows can expand to use thousands of simulations. Reducing the cost of each simulation could permit researchers to incorporate more physics in their models or expand their parameter searches. In addition, we aim to improve the reliability of these simulations by developing tools to automatically assess the validity of the simulations with regard to their specified simulation domains and numerical resolution. These tools would save time spent in post-hoc verification of results. The work presented in

this thesis is a preliminary step toward these broader objectives.

The rest of the thesis is organized in the following manner. In Chapter 2, we derive the electrostatic gyrokinetic system from the Fokker-Planck equation, the fundamental kinetic description of plasmas. Next, we derive the 1-D Vlasov-Poisson system from the same starting point. We briefly compare the two systems. Their shared features motivate our choice of the 1-D Vlasov-Poisson system as a test problem for closure modeling.

In Chapter 3, we construct our numerical scheme for solving the gyrokinetic and Vlasov-Poisson systems. We provide an overview of numerical methods and support our decision to use an Eulerian pseudo-spectral method to solve the equations. We present a detailed derivation of the projection of the systems of equations onto sets of orthogonal polynomial basis functions.

In Chapter 4, we introduce a machine learning closure model that reduces velocity-space resolution requirements for the 1-D Vlasov-Poisson system. We give a short overview of applications of machine learning to problems in plasma physics. Next, we provide details of our machine learning model of choice, reservoir computing, and we give context for its use cases. Then, we present simulations of the Vlasov-Poisson system and apply our machine learning closure model to it.

Chapter 5 studies the use of correlation functions, heat flux and energy spectra, and nonlinear energy transfer diagnostics to assess the validity of gyrokinetic turbulence simulations. We use the `GX` code to perform radially local flux tube simulations of electrostatic gyrokinetic turbulence in two model geometries, a tokamak case and a stellarator case. The tokamak case is a modification of a reference case often used to cross-verify gyrokinetic turbulence codes. The stellarator case simulates turbulence around a field line from a flux surface in the core of the Wendelstein 7-X stellarator. We highlight thresholds and trends in correlation length and resolution. Finally, we

summarize the main results of the thesis in Chapter 6. We conclude with a discussion of possible future extensions of the work in the previous chapters.

# Chapter 2: Theoretical Background: Kinetic Theory

The content in Section 2.3 of this chapter is adapted from the work published in Barbour et al. [2025], conducted in collaboration with Prof. William Dorland, Dr. Ian G. Abel, and Dr. Matt Landreman and has been reused with the permission of the surviving authors.

In this chapter, we derive the systems of equations necessary to form the theoretical basis for the results in this thesis. Though turbulence in fusion plasmas is our main motivation, we broaden the scope of this work to include the Vlasov-Poisson system, a simpler, one-dimensional model that shares some features of the electrostatic gyrokinetic system. We begin with the fundamental description of kinetic plasmas, the Fokker-Planck equation. We then derive the gyrokinetic equation in the electrostatic limit, in accordance with canonical approaches. Finally, we derive the Vlasov-Poisson system and elucidate its connection to electrostatic gyrokinetics. For simplicity, we elect to use Gaussian units throughout this thesis.

## 2.1 Defining our plasmas

In this thesis, we study fully ionized plasmas and neglect the influence of collisions with neutral particles. We also narrow our focus to weakly coupled plasmas, where the kinetic energy of the particles is significantly greater than their potential energy:

$$Z_s^2 e^2 / a T_s \ll 1, \tag{2.1}$$

where $Z_s e$ is the charge of particle species $s$, $T_s$ is its temperature in units of energy, and $a$ is the average interparticle spacing [Baalrud and Daligault, 2016]. The weak coupling condition

results in the Debye screening phenomenon, which diminishes the impact of Coulomb interactions between individual particles. The plasma exhibits collective behavior, with densely populated neighborhoods defined by the Debye length of the electrons,

$$n_e \lambda_{\mathrm{D}e}^3 \gg 1, \tag{2.2}$$

where the Debye length is defined as

$$\lambda_{\mathrm{D}e} \equiv \sqrt{\frac{T_e}{4\pi n_e e^2}}. \tag{2.3}$$

In addition, the plasmas we study in this thesis are non-relativistic, meaning that

$$v_{t,s}^2/c^2 \ll 1 \tag{2.4}$$

for all species, where we define the thermal speed of species $s$ as

$$v_{t,s} \equiv \sqrt{T_s/m_s}, \tag{2.5}$$

and $T_s$ is its equilibrium temperature in units of energy. In Section 2.2.4, we will precisely define the nature of the equilibrium and will explain why we wish to consider near-equilibrium plasmas. Finally, we also localize perturbations to scales shorter than the electron Debye length with the condition

$$k_\perp^2 \lambda_{\mathrm{D}e}^2 \ll 1, \tag{2.6}$$

where $k_\perp$ is the characteristic perpendicular wavenumber of a perturbation.

The weak coupling condition permits us to begin with the Fokker-Planck equation [Rosenbluth et al., 1957], which is the fundamental description of the time evolution of the statistical distribution function of a kinetic plasma species:

$$\frac{\mathrm{d}f_s}{\mathrm{d}t} = \frac{\partial f_s}{\partial t} + \mathbf{v}\cdot\nabla f_s + \frac{Z_s e}{m_s}\left(\widetilde{\mathbf{E}} + \frac{\mathbf{v}}{c}\times\widetilde{\mathbf{B}}\right)\cdot\frac{\partial f_s}{\partial \mathbf{v}} = C[f_s], \tag{2.7}$$

where $f_s(\mathbf{r},\mathbf{v},t)$ is the full distribution function of species $s$, with mass $m_s$, and charge $Z_s$ in units of the elementary charge, $e$. The total electric and magnetic fields are $\widetilde{\mathbf{E}}$ and $\widetilde{\mathbf{B}}$, respectively. This equation represents the effects of particle collisions with the Landau collision operator, $C[f_s]$, which describes the impact on particle species $s$ of collisions with all particle species [Landau, 1936]. The electric and magnetic fields are described by Maxwell's equations:

$$\nabla\cdot\widetilde{\mathbf{E}} = 4\pi\rho, \tag{2.8}$$

$$\nabla\cdot\widetilde{\mathbf{B}} = 0, \tag{2.9}$$

$$\nabla\times\widetilde{\mathbf{E}} = -\frac{1}{c}\frac{\partial\widetilde{\mathbf{B}}}{\partial t}, \tag{2.10}$$

$$\nabla\times\widetilde{\mathbf{B}} = \frac{1}{c}\left(4\pi\mathbf{j} + \frac{\partial\widetilde{\mathbf{E}}}{\partial t}\right). \tag{2.11}$$

The charge density $\rho$ and current $\mathbf{j}$ are given by

$$\rho = \sum_s Z_s e\int \mathrm{d}^3\mathbf{v}\, f_s, \tag{2.12}$$

and

$$\mathbf{j} = \sum_s Z_s e \int \mathrm{d}^3\mathbf{v}\mathbf{v} f_s, \tag{2.13}$$

respectively. Finally, the plasma obeys the quasineutrality condition:

$$\sum_s Z_s e \int \mathrm{d}^3\mathbf{v} f_s = 0, \tag{2.14}$$

which is a consequence of the Debye length being much smaller than the total system size.

As discussed in Chapter 1, the significant number of particles ($\sim 10^{20}$) in fusion plasmas preclude the possibility of a fast direct numerical solution to Newton's equations using currently available computational hardware. Similarly, the vast separation in spatial and temporal scales between microscopic turbulence and large-scale transport makes the direct simulation of the fundamental kinetic equation a challenging endeavor to pursue. To make progress, we can consider two limiting cases of (2.7), each with different levels of physics fidelity.

## 2.2 Electrostatic Gyrokinetic System with Adiabatic Electrons

One limiting case of the Fokker-Planck equation is the electrostatic gyrokinetic system. As a further simplification, we prescribe an adiabatic response for the electron distribution function, which eliminates the need to resolve electron-scale dynamics. This section follows the work of Catto [1978]; Antonsen and Lane [1980]; Frieman and Chen [1982]; Howes et al. [2006]; Abel et al. [2013]. The derivation follows Appendix A of [Howes et al., 2006] with a spatially varying mean distribution function and Sections 3–7 of [Abel et al., 2013] with no mean flow and no magnetic field perturbations.

### 2.2.1 Gyrokinetic orderings

The gyrokinetic equation is the result of applying a set asymptotic orderings and simplifying assumptions to (2.7). In addition to the assumptions that lead to the Fokker-Planck equation, the core additional assumptions that support the gyrokinetic equations are strong magnetic fields, small-amplitude perturbations, and low frequencies. In turn, these assumptions permit us to exploit the strong separation in spatial and temporal scales to split the distribution function into its mean and fluctuating components. Furthermore, the strong magnetic fields introduce a dominant anisotropy to the dynamics. When paired with the low frequencies, this permits us to average over the particle gyrations and reduce the domain to a (5+1)-D phase space characterized by three spatial dimensions, velocities parallel and perpendicular to the magnetic field, and time. To make this discussion more concrete, we define the gyrokinetic ordering assumptions in accordance with those employed by Howes et al. [2006]. First, we shall introduce the small expansion parameter $\epsilon$:

$$\epsilon = \frac{\rho_i}{a_N} \ll 1, \tag{2.15}$$

where we have introduced the typical ion Larmor radius,

$$\rho_i \equiv \frac{v_{t,i}}{\Omega_i}, \tag{2.16}$$

the ion cyclotron frequency,

$$\Omega_i \equiv \frac{Z_i e B}{m_i c}, \tag{2.17}$$

and the macroscopic normalizing length scale, $a_N$. Note that the asymptotic expansion parameter in (2.15) implies that the normalizing length, $a_N$, does not need a precise definition so long as the condition remains satisfied. Some examples of possible choices for $a_N$ include the minor radius of a tokamak or stellarator and several multiples of the typical parallel wavelength of turbulent fluctuations in an astrophysical plasma. This parameter embeds the strong magnetization into the system by ensuring that the ion Larmor radius is small relative to macroscopic system parameters.

Next, we consider the spatial ordering that arises from the strong magnetization condition. Strong magnetization causes charged particles to The strong magnetization promotes a natural *separation of scales* such that the perpendicular length scales remain smaller than the parallel length scales, which leads to the ordering:

$$\frac{k_\parallel}{k_\perp} \sim \mathcal{O}(\epsilon), \tag{2.18}$$

where $\mathbf{k} = k_\parallel \widehat{\mathbf{b}} + \mathbf{k}_\perp$ represents the wavenumber of a perturbation, where we have introduced the unit vector in the direction of the mean magnetic field:

$$\widehat{\mathbf{b}} \equiv \frac{\mathbf{B}}{|\mathbf{B}|} = \frac{\mathbf{B}}{B}, \tag{2.19}$$

and $\mathbf{B}$ is the mean magnetic field. We shall explore this more in Chapter 5 through a study of the perpendicular and parallel correlation lengths.

With the expansion parameter in hand, we can proceed to write explicit definitions for the other orderings. The low frequency condition asserts that the frequencies, $\omega$, of the turbulent

fluctuations are slower than the ion cyclotron frequency:

$$\frac{\omega}{\Omega_i} \sim \frac{v_{t,i}}{a_N}\frac{1}{\Omega_i} \sim \mathcal{O}(\epsilon). \tag{2.20}$$

In the time that an ion experiences a turbulent fluctuation, it undergoes many cyclotron orbits around a magnetic field line. This separation of time scales permits us to average over particle orbits and consider the time evolution of the distribution function in terms of charged rings in guiding-center coordinates.

We formally split the distribution function and fields into their mean and fluctuating components:

$$f_s = F_s + \delta f_s, \tag{2.21}$$

$$\widetilde{\mathbf{E}} = \mathbf{E} + \delta\mathbf{E}, \tag{2.22}$$

$$\widetilde{\Phi} = \Phi + \delta\Phi, \tag{2.23}$$

$$\widetilde{\mathbf{B}} = \mathbf{B} + \delta\mathbf{B}. \tag{2.24}$$

The premise that the fluctuations can be studied separately from the mean distribution imposes an additional temporal ordering assumption. The timescales of the fluctuations must be well-separated from the "transport timescales" on which the mean background fields evolve. A canonical ordering expression for this condition [Frieman and Chen, 1982; Howes et al., 2006] is

$$\frac{1}{\tau_{\text{transport}}} \equiv \frac{1}{F_{0s}}\frac{\partial F_{0s}}{\partial t} \sim \mathcal{O}(\epsilon^3), \tag{2.25}$$

where $F_{0s}$ is the equilibrium distribution function, and $\tau_{\rm transport}$ is the macroscopic transport timescale. A more detailed discussion can be found in Callen et al. [2010].

We assume that the electric and magnetic fields follow the ordering

$$\frac{|\widetilde{\mathbf{E}}|}{|\widetilde{\mathbf{B}}|} \sim \frac{v_{t,i}}{c}, \tag{2.26}$$

which establishes that the magnetic field dominates the electric field. This is equivalent to assuming that the $\mathbf{E} \times \mathbf{B}$ flows are non-relativistic [Abel et al., 2013]. In addition, the perturbations must be small relative to the background:

$$\frac{\delta f_s}{F_s} \sim \mathcal{O}(\epsilon). \tag{2.27}$$

Note that in the electrostatic limit, we forbid any time-dependent perturbations to the magnetic field. Faraday's law yields a rotationless electric field on the timescales of the turbulent fluctuations in this limit:

$$\nabla \times \widetilde{\mathbf{E}} = -\frac{1}{c}\frac{\partial \widetilde{\mathbf{B}}}{\partial t} = 0. \tag{2.28}$$

This also permits us to use the familiar electrostatic potential formulation of the electric field:

$$\widetilde{\mathbf{E}} = -\nabla \widetilde{\Phi}, \tag{2.29}$$

where $\Phi(\mathbf{r})$ is a scalar function of particle position.

### 2.2.2 The Gyroaverage

A defining feature of the gyrokinetic system is that it represents the dynamics with a reduced model in a (5+1)-D phase space, while maintaining accuracy up to order $\mathcal{O}(\epsilon)$. The dimension reduction is possible due to the strong magnetic fields and the separation of timescales introduced in Section 2.2.1, which permits us to eliminate explicit dependence on the gyrophase of particle orbits by suitable averaging. In this section, we shall use guiding center coordinates following Catto [1978]. Particle positions, $\mathbf{r}$, are related to their guiding center positions $\mathbf{R}_s$ by the transformation

$$\mathbf{R}_s = \mathbf{r} + \frac{\mathbf{v} \times \widehat{\mathbf{b}}}{\Omega_s}. \tag{2.30}$$

The total velocity and its perpendicular and parallel components are related by

$$\mathbf{v} = v_\parallel \widehat{\mathbf{b}} + v_\perp (\cos\vartheta \widehat{\mathbf{e}_2} - \sin\vartheta \widehat{\mathbf{e}_1}), \tag{2.31}$$

where $\widehat{\mathbf{e}_1}$ and $\widehat{\mathbf{e}_2}$ are orthonormal unit vectors that define the plane perpendicular to the magnetic field such that $\widehat{\mathbf{b}} = \widehat{\mathbf{e}_2} \times \widehat{\mathbf{e}_1}$ and $\vartheta$ is the gyrophase.

We then introduce two gyroaveraging operators, $\langle\,\rangle_{\mathbf{R}_s}$ and $\langle\,\rangle_{\mathbf{r}_s}$, where the subscripts denote which position variable remains constant while computing the average. We define these operators by

$$\langle g(\mathbf{r}, \mathbf{v}, t)\rangle_{\mathbf{R}_s} \equiv \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\vartheta g\left(\mathbf{R}_s - \frac{\mathbf{v} \times \widehat{\mathbf{b}}}{\Omega_s}, \mathbf{v}, t\right), \tag{2.32}$$

and

$$\langle g(\mathbf{R}_s, \mathbf{v}, t)\rangle_{\mathbf{r}} \equiv \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\vartheta g\left(\mathbf{r} + \frac{\mathbf{v} \times \widehat{\mathbf{b}}}{\Omega_s}, \mathbf{v}, t\right). \tag{2.33}$$

As a first step toward deriving the gyrokinetic system, we shall express the Fokker-Planck equation (2.7) in the coordinates $(\mathbf{r}, v_\parallel, v_\perp, \vartheta, t)$:

$$\frac{\partial f_s}{\partial t} + v_\parallel \frac{\partial f_s}{\partial z} + \mathbf{v}_\perp \cdot \nabla_\perp f_s - \frac{Ze}{m_s}\frac{\partial \Phi}{\partial z}\frac{\partial f_s}{\partial v_\parallel} + \frac{Ze}{m_s}\left(-\nabla_\perp \Phi + \frac{\mathbf{v}_\perp \times \mathbf{B}}{c}\right) \cdot \frac{\partial f_s}{\partial \mathbf{v}_\perp} = C[f_s]. \quad (2.34)$$

For ease of explanation, we consider an imposed homogeneous magnetic field:

$$\mathbf{B} = B\widehat{\mathbf{b}} = B\widehat{\mathbf{z}}. \quad (2.35)$$

Using the small perturbation assumption (2.27), we can formally split the distribution function into its mean and fluctuating components, $F_s$ and $\delta f$, respectively:

$$f_s = F_s + \delta f_s(\mathbf{r}, v_\parallel, v_\perp, \theta). \quad (2.36)$$

We desire to enumerate the terms of the gyrokinetic equation order-by-order. First, we shall identify the terms of the kinetic equation at each order of the expansion parameter. Then, we will gyroaverage the resulting kinetic equation to obtain the electrostatic gyrokinetic equation. To facilitate this process, we first expand the mean and fluctuating portions of the distribution function as in the work of Abel et al. [2013]:

$$F_s = F_{0s} + F_{1s} + F_{2s} + \ldots, \quad (2.37)$$

$$\delta f_s = \delta f_{1s} + \delta f_{2s} + \ldots, \quad (2.38)$$

where each subscript indicates the order of the expansion:

$$F_{0s} \sim f_s, \tag{2.39}$$

$$F_{1s} \sim \delta f_{1s} \sim \epsilon f_s, \tag{2.40}$$

$$F_{2s} \sim \delta f_{2s} \sim \epsilon^2 f_s. \tag{2.41}$$

### 2.2.3 Zeroth order of the gyrokinetic equation

By the orderings in Section 2.2.1, the lowest order of the kinetic equation (2.34) is

$$\mathbf{v}_\perp \cdot \nabla_\perp F_{0s} + \frac{Z_s e}{m_s}\left(\frac{\mathbf{v}_\perp \times \mathbf{B}}{c}\right) \cdot \frac{\partial F_{0s}}{\partial \mathbf{v}_\perp} = 0. \tag{2.42}$$

Expressing this equation in terms of the guiding center velocity coordinates $(v_\parallel, v_\perp, \theta)$ will reveal an important property of $F_{0s}$. To recast the equation in terms of those coordinates, we must first solve (2.31) for those variables in terms of the velocities in the perpendicular plane:

$$v_\perp = \sqrt{v_1^2 + v_2^2}, \tag{2.43}$$

$$\vartheta = \arctan\left(\frac{v_1}{v_2}\right), \tag{2.44}$$

where

$$v_1 = \mathbf{v} \cdot \widehat{\mathbf{e}}_1 = -v_\perp \sin\vartheta, \tag{2.45}$$

$$v_2 = \mathbf{v} \cdot \widehat{\mathbf{e}}_2 = v_\perp \cos\vartheta. \tag{2.46}$$

Using these definitions we have

$$\frac{\mathbf{v}_\perp \times \mathbf{B}}{c} = \left(\frac{v_\perp B}{c}\right)\left(\cos\vartheta\widehat{\mathbf{e}}_1 + \sin\vartheta\widehat{\mathbf{e}}_2\right). \tag{2.47}$$

Similarly, we can express the gradient with respect to perpendicular velocity as

$$\frac{\partial}{\partial \mathbf{v}_\perp} = \frac{\partial}{\partial v_1}\widehat{\mathbf{e}}_1 + \frac{\partial}{\partial v_2}\widehat{\mathbf{e}}_2. \tag{2.48}$$

We can then use the chain rule and (2.43) to transform the velocity derivatives from coordinates $v_1$ and $v_2$ to $v_\perp$ and $\vartheta$. First,

$$\begin{aligned}
\frac{\partial}{\partial v_1} &= \frac{\partial}{\partial v_\perp}\frac{\partial v_\perp}{\partial v_1} + \frac{\partial}{\partial \vartheta}\frac{\partial \vartheta}{\partial v_1}, \\
&= \frac{v_1}{v_\perp}\frac{\partial}{\partial v_\perp} + \frac{v_2}{v_\perp^2}\frac{\partial}{\partial \vartheta}, \\
&= -\sin\vartheta\frac{\partial}{\partial v_\perp} + \frac{\cos\vartheta}{v_\perp}\frac{\partial}{\partial \vartheta}.
\end{aligned} \tag{2.49}$$

Similarly,

$$\begin{aligned}
\frac{\partial}{\partial v_2} &= \frac{\partial v_\perp}{\partial v_2}\frac{\partial}{\partial v_\perp} + \frac{\partial \vartheta}{\partial v_2}\frac{\partial}{\partial \vartheta}, \\
&= \frac{v_2}{v_\perp}\frac{\partial}{\partial v_\perp} - \frac{v_1}{v_\perp^2}\frac{\partial}{\partial \vartheta}, \\
&= \cos\vartheta\frac{\partial}{\partial v_\perp} + \frac{\sin\vartheta}{v_\perp}\frac{\partial}{\partial \vartheta}.
\end{aligned} \tag{2.50}$$

The scalar product of (2.47) and (2.48) then results in

$$\begin{aligned}\frac{\mathbf{v}_\perp \times \mathbf{B}}{c} \cdot \frac{\partial}{\partial \mathbf{v}_\perp} &= \left(\frac{v_\perp B}{c}\right) \cos\vartheta \left(-\sin\vartheta \frac{\partial}{\partial v_\perp} + \frac{\cos\vartheta}{v_\perp}\frac{\partial}{\partial \vartheta}\right) \\ &\quad + \left(\frac{v_\perp B}{c}\right) \sin\vartheta \left(\cos\vartheta \frac{\partial}{\partial v_\perp} + \frac{\sin\vartheta}{v_\perp}\frac{\partial}{\partial \vartheta}\right), \\ &= \left(\frac{v_\perp B}{c}\right)\left(\frac{\cos^2\vartheta}{v_\perp} + \frac{\sin^2\vartheta}{v_\perp}\right)\frac{\partial}{\partial \vartheta}, \\ &= \frac{B}{c}\frac{\partial}{\partial \vartheta}.\end{aligned} \tag{2.51}$$

Using this result, the zeroth order of the kinetic equation (2.42) then simplifies to

$$\begin{aligned}\mathbf{v}_\perp \cdot \nabla_\perp F_{0s} + \frac{Z_s e}{m_s}\left(\frac{\mathbf{v}_\perp \times \mathbf{B}}{c}\right) \cdot \frac{\partial F_{0s}}{\partial \mathbf{v}_\perp} &= 0, \\ \mathbf{v}_\perp \cdot \nabla_\perp F_{0s} + \frac{Z_s e B}{m_s c}\frac{\partial F_{0s}}{\partial \vartheta} &= 0, \\ \mathbf{v}_\perp \cdot \nabla_\perp F_{0s} + \Omega_s \frac{\partial F_{0s}}{\partial \vartheta} &= 0.\end{aligned} \tag{2.52}$$

We can show that the left-hand side simplifies to a derivative in gyrophase at constant guiding center position by using (2.30), (2.31), and (2.51):

$$\begin{aligned}\Omega_s \frac{\partial F_{0s}}{\partial \vartheta}\bigg|_{\mathbf{R}_s} &= \Omega_s \left[\left(\frac{\partial \mathbf{r}}{\partial \vartheta}\right)\bigg|_{\mathbf{R}_s} \cdot \nabla F_{0s} + \left(\frac{\partial \mathbf{v}}{\partial \vartheta}\right)\bigg|_{\mathbf{R}_s} \cdot \frac{\partial F_{0s}}{\partial \mathbf{v}}\right] \\ &= \Omega_s \left[-\frac{\mathbf{v}_\perp}{\Omega_s} \cdot \nabla F_{0s} - (\mathbf{v}_\perp \times \widehat{\mathbf{z}}) \cdot \frac{\partial F_{0s}}{\partial \mathbf{v}}\right] \\ &= -\Omega_s \left[\frac{\mathbf{v}_\perp}{\Omega_s} \cdot \nabla F_{0s} + \frac{\partial F_{0s}}{\partial \vartheta}\bigg|_{\mathbf{r}}\right].\end{aligned} \tag{2.53}$$

Applying this result to (2.52) yields

$$-\Omega_s \frac{\partial F_{0s}}{\partial \vartheta}\bigg|_{\mathbf{R}_s} = 0. \tag{2.54}$$

The immediate consequence of this equation is that the mean distribution function $F_{0s}$ is independent of the gyrophase to lowest order. This observation aligns with our low-frequency ordering assumption, $\omega \ll \Omega_i$. It also introduces a natural reference point for higher orders in the expansion parameter. The largest terms captured by the kinetic equation for species $s$ are of order $\mathcal{O}(\Omega_s F_{0s})$. All other terms shall be compared to this reference.

### 2.2.4 First order of the kinetic equation

We now derive the first order of the kinetic equation. To this order, (2.34) becomes

$$v_\parallel \frac{\partial F_{0s}}{\partial z} + \mathbf{v}_\perp \cdot \nabla_\perp (F_1 + \delta f_1) - \frac{Z_s e}{m_s} \nabla_\perp \Phi \cdot \frac{\partial F_{0s}}{\partial \mathbf{v}_\perp} + \Omega_s \frac{\partial}{\partial \vartheta} (F_1 + \delta f_1) = C[F_{0s}, F_{0s}]. \tag{2.55}$$

Note that the $k_\parallel / k_\perp \sim \epsilon$ ordering separates $(\partial/\partial z)$ and $\nabla_\perp$, and we have also used the result from (2.52). At this stage, the distribution function is unknown. We can transform this equation to eliminate $F_1$ and $\delta f_1$ from the equation, which will grant us enough information to specify $F_{0s}$. We make two additional assumptions:

1. Spatial averages of perturbed quantities are zero.

2. The fluctuations are statistically decorrelated over long distances in the domain such that we can assume periodic boundary conditions.

Under these assumptions, the spatial average of the left-hand-side of (2.55) is 0. We can then multiply (2.55) by $1 + \ln F_{0s}$ and integrate over phase space to obtain

$$\int \mathrm{d}x^3 \int \mathrm{d}v^3 \, \ln F_{0s} C[F_{0s}, F_{0s}] = 0. \tag{2.56}$$

This equation shows that the Landau collision operator satisfies the Boltzmann H-theorem [Reif, 1965], which states that the kinetic entropy of the system cannot decrease due to collisions. We can first consider the integral over velocity space in (2.56), which is known to be related to the expression for the rate of entropy production due to self-collisions [Helander and Sigmar, 2002]:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\int \mathrm{d}v^3 \ln F_{0s} C[F_{0s}, F_{0s}] \geq 0. \tag{2.57}$$

This inequality holds for all neighborhoods in space, as the local, linearized form of the Landau collision operator also satisfies (2.57). The strict equality in the integral equation (2.56) is uniquely satisfied by the maximal entropy solution, which is local Maxwellian velocity distribution [Helander and Sigmar, 2002]:

$$F_{0s} = n_s \left(\frac{m_s}{2\pi T_s}\right)^{3/2} \exp\left[-\frac{m_s(v_\perp^2 + v_\parallel^2)}{2T_s}\right], \tag{2.58}$$

where $n_s(\mathbf{r})$ and $T_s(\mathbf{r})$ are the density and temperature of plasma species $s$.

Recalling that $(\partial F_{0s}/\partial \vartheta) = 0$ and using the expressions for the velocity derivatives (2.49) and (2.50), we can evaluate the velocity gradient of the mean distribution function:

$$\begin{aligned} \frac{\partial F_{0s}}{\partial \mathbf{v}} &= \frac{\partial}{\partial \mathbf{v}}\left[n_s \left(\frac{m_s}{2\pi T_s}\right)^{3/2} \exp\left[-\frac{m_s|\mathbf{v}|^2}{2T_s}\right]\right] \\ &= -\frac{m}{T_s} F_{0s}\mathbf{v}. \end{aligned} \tag{2.59}$$

Substituting this explicit expression for the velocity gradient (2.59) into the first-order kinetic

equation (2.55) yields

$$\begin{aligned} v_\parallel \frac{\partial F_{0s}}{\partial z} + \mathbf{v}_\perp \cdot \nabla_\perp (F_1 + \delta f_1) + \Omega_s \frac{\partial}{\partial \vartheta}(F_1 + \delta f_1) &= -\frac{Z_s e}{m_s} \nabla_\perp \Phi \cdot \left(\frac{m_s}{T_s}\right) F_{0s} \mathbf{v}_\perp \\ &= -\frac{Z_s e}{T_s} \nabla_\perp \Phi \cdot F_{0s} \mathbf{v}_\perp. \end{aligned} \tag{2.60}$$

If we gyroaverage this order of the expansion at constant guiding center position $\mathbf{R}_s$, the first term on the left-hand side vanishes. By (2.53), the second and third terms combine to become a partial derivative in $\vartheta$:

$$-\Omega_s \left( \left.\frac{\partial F_{1s}}{\partial \vartheta}\right|_{\mathbf{R}_s, v_\parallel, v_\perp} + \left.\frac{\partial \delta f_{1s}}{\partial \vartheta}\right|_{\mathbf{R}_s, v_\parallel, v_\perp} \right) = -\frac{Z_s e}{T_s} \nabla_\perp \Phi \cdot \mathbf{v}_\perp F_{0s}. \tag{2.61}$$

Averaging over the perturbed quantities demonstrates that $F_{1s}$ must also be independent of the gyrophase:

$$-\Omega_s \left.\frac{\partial F_{1s}}{\partial \vartheta}\right|_{\mathbf{R}_s, v_\parallel, v_\perp} = 0. \tag{2.62}$$

Now, we can determine the fluctuating portion of the distribution function at this order, $\delta f_{1s}$, by evaluating the fluctuating part of (2.61) :

$$\begin{aligned} -\Omega_s \left.\frac{\partial \delta f_{1s}}{\partial \vartheta}\right|_{\mathbf{R}_s, v_\parallel, v_\perp} &= -\frac{Z_s e}{T_s} \nabla_\perp \Phi \cdot \mathbf{v}_\perp F_{0s} \\ &= \Omega_s \frac{Z_s e}{T_s} \left.\frac{\partial \Phi}{\partial \vartheta}\right|_{\mathbf{R}_s, v_\parallel, v_\perp} F_{0s}. \end{aligned} \tag{2.63}$$

Integrating this equation with respect to $\vartheta$ yields the solution:

$$\delta f_{1s} = -\frac{Z_s e \Phi}{T_s} F_{0s} + h_s(\mathbf{R}_s, v_\parallel, v_\perp, t), \tag{2.64}$$

where we have introduced the gyrophase-independent distribution function of the perturbed rings of charge, $h_s$. Note that the Boltzmann response that the first term represents still depends on the gyrophase, as $\Phi$ is a function of particle position. Collecting these results yields this solution for the distribution function:

$$f_s = F_{0s} - \frac{Z_s e\Phi}{T_s} F_{0s} + F_{1s} + h_s + F_{2s} + \delta f_{2s} + \mathcal{O}(\epsilon^3 f), \tag{2.65}$$

where we know the functional form of the zeroth order of the mean distribution function:

$$F_{0s} = n_s \left(\frac{m_s}{2\pi T_s}\right)^{3/2} \exp\left[-\frac{m_s(v_\perp^2 + v_\parallel^2)}{2T_s}\right]. \tag{2.66}$$

In the next section, we will seek to determine the evolution equation that governs the small scale fluctuations, $h_s$. To close the system, we will need to gain information about the terms of order $\mathcal{O}(\epsilon^2)$, $F_{2s}$ and $\delta f_{2s}$. We will then gyroaverage the resulting evolution equation to arrive at the gyrokinetic equation in the electrostatic limit.

### 2.2.5 Second order of the gyrokinetic equation

First, we will transform the kinetic equation (2.34) to guiding-center coordinates, $(\mathbf{R}_s, v_\parallel, v_\perp, \vartheta)$:

$$\frac{\partial f_s}{\partial t} + \frac{\mathrm{d}\mathbf{R}_s}{\mathrm{d}t} \cdot \frac{\partial f_s}{\partial \mathbf{R}_s} + \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} \cdot \frac{\partial f_s}{\partial \mathbf{v}} = C[f_s]. \tag{2.67}$$

We now need an explicit expression for $(\mathrm{d}\mathbf{R}_s/\mathrm{d}t)$. To derive this expression we use the velocity and acceleration of the charged particles:

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \mathbf{v}, \tag{2.68}$$

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \frac{Z_s e}{m_s}\left(\mathbf{E} + \frac{\mathbf{v}\times\mathbf{B}}{c}\right). \tag{2.69}$$

With these definitions and (2.30), we can now derive the time derivative of the guiding center position:

$$\begin{aligned}
\frac{\mathrm{d}\mathbf{R}_s}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{r} + \frac{\mathbf{v}\times\widehat{\mathbf{z}}}{\Omega_s}\right) \\
&= \mathbf{v} + \frac{1}{\Omega_s}\left(\frac{Z_s e}{m_s}\right)\left(\mathbf{E} + \frac{\mathbf{v}\times\mathbf{B}}{c}\right)\times\widehat{\mathbf{z}} \\
&= \mathbf{v} + \frac{c}{B}\mathbf{E}\times\widehat{\mathbf{z}} + (\mathbf{v}\times\widehat{\mathbf{z}}\times\widehat{\mathbf{z}}) \\
&= \mathbf{v} - \frac{c}{B}\nabla\Phi\times\widehat{\mathbf{z}} - \widehat{\mathbf{z}}\times(\mathbf{v}\times\widehat{\mathbf{z}}) \\
&= \mathbf{v} - \frac{c}{B}\nabla\Phi\times\widehat{\mathbf{z}} - [\mathbf{v}(\widehat{\mathbf{z}}\cdot\widehat{\mathbf{z}}) - \widehat{\mathbf{z}}(\widehat{\mathbf{z}}\cdot\mathbf{v})] \\
&= v_\parallel\widehat{\mathbf{z}} - \frac{c}{B}\nabla_\perp\Phi\times\widehat{\mathbf{z}}.
\end{aligned} \tag{2.70}$$

After substituting this final expression into (2.67), we find

$$\frac{\partial f_s}{\partial t} + v_\parallel\frac{\partial f_s}{\partial z} - \frac{c}{B}(\nabla\Phi\times\widehat{\mathbf{z}})\cdot\frac{\partial f_s}{\partial\mathbf{R}_s} + \frac{Z_s e}{m_s}\left(-\nabla_\perp\Phi + \frac{\mathbf{v}\times\mathbf{B}}{c}\right)\cdot\frac{\partial f_s}{\partial\mathbf{v}} = C[f_s]. \tag{2.71}$$

Similarly to the previous sections, we shall now isolate the second-order terms in (2.71) by using our derived functional forms of the previous orders of the kinetic equation (2.64–2.66).

$$\frac{\partial h_s}{\partial t} + \left[v_\parallel \widehat{\mathbf{z}} - \frac{c}{B}(\nabla_\perp \Phi \times \widehat{\mathbf{z}})\right] \cdot \frac{\partial}{\partial \mathbf{R}_s}(F_{1s} + h_s) - \frac{Z_s e}{m_s}\nabla_\perp \Phi \cdot \left(\frac{\mathbf{v}_\perp}{v_\perp}\frac{\partial}{\partial v_\perp}(F_{1s} + h)\right)$$
$$= \frac{\partial}{\partial t}\left(\frac{Z_s e \Phi}{T_s} F_{0s}\right) + \frac{Z_s e \Phi}{T_s} v_\parallel \frac{\partial F_{0s}}{\partial z} + \frac{Z_s^2 e^2}{T_s^2}\Phi \nabla_\perp \Phi \cdot F_{0s}\mathbf{v}_\perp - \Omega_s \frac{\partial}{\partial \vartheta}(F_2 + \delta f_{2s}) - C[F_{0s}, h] - C[h, F_{0s}]. \tag{2.72}$$

First, we define the second term of (2.70) as the $\mathbf{E} \times \mathbf{B}$ velocity, and we also introduce its gyroaverage:

$$\mathbf{v}_E = \frac{c}{B}\widehat{\mathbf{z}} \times \nabla_\perp \Phi, \qquad \langle \mathbf{v}_E \rangle_{\mathbf{R}} = \frac{c}{B}\left(\widehat{\mathbf{z}} \times \frac{\partial \langle \Phi \rangle_{\mathbf{R}}}{\partial \mathbf{R}_s}\right). \tag{2.73}$$

Observe that the unknown second-order terms $\delta f_{2s}$ and $F_{2s}$ appear only in derivatives of the gyrophase. By construction, these functions must be periodic in $\vartheta$, which implies that the integrals of their $\vartheta$ derivatives vanish over one orbit. Thus, we can gyroaverage this equation at fixed guiding center position by using (2.32), which eliminates $\delta f_{2s}$ and $F_{2s}$. As before, we can split the evolution equation for this order into the mean and fluctuating components. The gyroaverage of the fluctuating portion of (2.72) becomes the electrostatic, second-order gyrokinetic equation in straight, homogeneous magnetic field-line geometry:

$$\frac{\partial h_s}{\partial t} + \left[v_\parallel \widehat{\mathbf{z}} + \langle \mathbf{v}_E \rangle_{\mathbf{R}}\right] \cdot \frac{\partial}{\partial \mathbf{R}_s} h_s = \frac{\partial}{\partial t}\left(\frac{Z_s e}{T_s} F_{0s} \langle \Phi \rangle_{\mathbf{R}_s}\right) + \langle C[h_s] \rangle_{\mathbf{R}_s}. \tag{2.74}$$

### 2.2.6 Quasineutrality: Determining $\Phi$

Finally, we use the fluctuating component of the quasineutrality condition (2.14) to order $\mathcal{O}(\epsilon)$ to determine $\langle\Phi\rangle_{\mathbf{R}_s}$, allowing us to close (2.74):

$$\sum_s Z_s e \int \mathrm{d}^3\mathbf{v}\langle h_s\rangle_{\mathbf{r}} = \sum_s \frac{Z_s^2 e^2 n_s}{T_s}\Phi. \tag{2.75}$$

As mentioned previously, we use the electrostatic approximation and neglect time-dependent perturbations to the magnetic field. We will also leverage the vast difference in mass between electrons and ions to make an additional simplifying assumption: the adiabatic electron approximation. Consider the linearized form of the gyrokinetic equation without collisions (2.74):

$$\begin{aligned} \mathrm{i}\omega h_s + \mathrm{i}k_\parallel v_\parallel &= \mathrm{i}\omega\left(\frac{Z_s e}{T_s}F_{0s}\langle\Phi\rangle_{\mathbf{R}}\right) \\ h_s &= \frac{\omega}{\omega + k_\parallel v_\parallel}\left(\frac{Z_s e}{T_s}F_{0s}\langle\Phi\rangle_{\mathbf{R}}\right). \end{aligned} \tag{2.76}$$

When evaluating this linearized equation for $h_s$ for either electrons or ions, all terms of the equation are of similar orders to each other with the exception of the $k_\parallel v_\parallel$ term in the divisor. We assume that $v_\parallel \sim v_{ts}$. For $T_e \approx T_i$, the ratio between the thermal speed of an electron and the thermal speed of an ion is approximately

$$\frac{v_{te}}{v_{ti}} \sim \sqrt{\frac{m_i}{m_e}}. \tag{2.77}$$

The ratio between the mass of a proton and the mass of an electron is roughly 1,836 [NIST, 2024]. Therefore, the electrons move much more quickly than the ions. Additionally, the relationship between the ion and electron thermal speeds with the linearized gyrokinetic equation implies that

$h_e \ll h_i$. We can exploit this to make the further simplifying assumption that in addition to the rapid, mean thermalization toward a Maxwellian velocity distribution, spatiotemporal perturbations to the electron distribution function are prescribed by the Boltzmann response to variations in the electrostatic potential:

$$f_e \approx F_{0e} + \frac{e}{T_e}\Phi. \tag{2.78}$$

This approximation permits us to neglect kinetic dynamics at the electron scales, which are shorter and faster than the ion scales. In the limit of a single ion species, the plasma then evolves according to the gyrokinetic equation for the ions (2.74), and the quasineutrality condition (2.75) reduces to

$$Z_i e \int \mathrm{d}^3\mathbf{v} \langle h \rangle_{\mathbf{r}} = n_0 e^2 \left[ \Phi \left( \frac{Z_i^2}{T_i} \right) + \left( \frac{1}{T_e} \right) \left[ \Phi - \langle\langle \Phi \rangle\rangle \right] \right], \tag{2.79}$$

where we have introduced the flux-surface average of the potential, $\langle\langle \Phi \rangle\rangle$. In the gyrokinetic turbulence studies in Chapter 5, we leverage the theoretical and empirical evidence for the existence of stable magnetohydrodynamic (MHD) force balance solutions that result in surfaces of constant magnetic flux that are crucial to magnetic confinement fusion geometries. The flux-surface average term ensures that the mean of the perturbations in the electron density is zero over the flux surface [Dorland and Hammett, 1993]. The work of Abel and Cowley [2013] demonstrates that this term arises from first-principles in an rigorous collisional electrostatic limit of gyrokinetics.

## 2.3 1-D Vlasov-Poisson System

As our test problem for the machine-learning closure model we implement in Chapter 4, we study the dynamics of a one-dimensional, collisionless, electrostatic hydrogen plasma. We focus

on the electron dynamics, treating the ions as a cold, neutralizing background. The equations that describe this plasma are the 1-D, 1-V Vlasov-Poisson system [Vlasov, 1968]. In this section, we begin with the Fokker-Planck equation, derive the Vlasov-Poisson system from the appropriate limit, and normalize the system for the purpose of simulation. The process is simple in comparison to the derivation of the electrostatic gyrokinetic system, but it will be instructive to compare their features.

In the absence of collisions, (2.7) becomes the Vlasov equation and represents conservation of phase-space volume [Nicholson, 1983]:

$$\frac{\mathrm{d}f_s}{\mathrm{d}t} = \frac{\partial f_s}{\partial t} + \mathbf{v}\cdot\nabla f_s + \frac{Z_s e}{m_s}\left(\mathbf{E} + \frac{\mathbf{v}}{c}\times\mathbf{B}\right)\cdot\frac{\partial f_s}{\partial \mathbf{v}} = 0. \tag{2.80}$$

Three simplifying assumptions are necessary to proceed.

1. There is no magnetic field.

2. Neglect perpendicular dynamics.

3. The ions are a cold, neutralizing background.

After these assumptions, all that remains of the Fokker-Planck equation is an evolution equation for an electron plasma, the 1-D Vlasov equation:

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial z} - \frac{eE}{m_e}\frac{\partial f}{\partial v} = 0, \tag{2.81}$$

where $f(z, v, t)$ is the electron distribution function, $e$ is the elementary charge, and $m_e$ is the mass of an electron. $E(z, t)$ is the electric field, which we calculate using Poisson's equation,

$$-\frac{\partial^2 \Phi}{\partial z^2} = 4\pi\rho, \quad E = -\frac{\partial \Phi}{\partial z}, \tag{2.82}$$

where $\rho(z,t)$ is the total electric charge density, including both ions and electrons, and $\Phi(z,t)$ is the electric potential. For this quasi-neutral plasma with cold ions, we can evaluate the charge density as

$$\rho(z,t) = e\left(n_0 - \int \mathrm{d}v f\right), \tag{2.83}$$

where $n_0$ is the mean number density of both species.

Plasmas with near-Maxwellian velocity distributions are of particular interest for magnetic fusion applications. Therefore, we separate the distribution function into a Maxwellian mean and time-dependent perturbations:

$$f(z,v,t) = F_0(v) + g(z,v,t), \tag{2.84}$$

where

$$F_0(v) = \left(\frac{n_0}{v_{te}}\right)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-v^2/2v_{te}^2}, \tag{2.85}$$

is the one-dimensional Maxwellian velocity distribution for the electrons, and $g(z,v,t)$ is the fluctuating part of the distribution function. We have defined the electron thermal speed as $v_{te} = \sqrt{T_e/m_e}$, where $T_e$ is the temperature of the electrons in units of energy. The Vlasov equation supports longitudinal Langmuir waves, whose dynamics occur on time scales on the order of the plasma frequency,

$$\omega_{pe} \equiv \sqrt{\frac{4\pi n_0 e^2}{m_e}}. \tag{2.86}$$

The characteristic length-scale for those oscillations is the Debye length,

$$\lambda_D \equiv \sqrt{\frac{T_e}{4\pi n_0 e^2}}. \tag{2.87}$$

To support numerical integration of this system of equations, we normalize the variables into dimensionless forms, denoted with primes:

$$\begin{aligned} &t \equiv \frac{1}{\omega_{pe}} t', \quad v \equiv v_{te} v', \quad z \equiv \lambda_D z', \quad E \equiv \frac{T_e}{e\lambda_D} E', \quad f \equiv F_0 + g \equiv \frac{n_0}{v_t}(F'_M + g'), \\ &\rho \equiv n_0 e \rho', \quad \Phi \equiv \frac{T_e}{e}\Phi', \quad F'_M \equiv \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-v'^2/2}, \quad \int \mathrm{d}v' F'_M = 1. \end{aligned} \tag{2.88}$$

After dropping the primes for legibility, we can now write the model equations in dimensionless form:

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial z} - E\frac{\partial f}{\partial v} = 0. \tag{2.89}$$

$$\frac{\partial E}{\partial z} = 1 - \int \mathrm{d}v f. \tag{2.90}$$

Finally, after splitting the distribution function into its mean and fluctuating components, we arrive at the normalized evolution equation for the fluctuating part of the electron distribution function:

$$\frac{\partial g}{\partial t} + v\frac{\partial g}{\partial z} + vEF_M - E\frac{\partial g}{\partial v} = 0, \tag{2.91}$$

$$\frac{\partial E}{\partial z} = -\int \mathrm{d}v\, g. \tag{2.92}$$

The Vlasov-Poisson system shares some features with the electrostatic gyrokinetic system

that allow it to be a useful test model for the closure problem. The gyrokinetic parallel streaming term, $v\nabla_{\parallel}h$, is a generalization of the free streaming term in 2.91, $v(\partial g)/(\partial z)$. Both equations have a quadratic nonlinear term. In gyrokinetics, the $\mathbf{E} \times \mathbf{B}$ drift is the quadratic nonlinearity. In the 1-D Vlasov equation, the nonlinearity is a parallel velocity derivative term, $E(\partial g)/(\partial v)$. The equivalent term is at the next order in the gyrokinetic equation, so it is not present. The strong magnetization in gyrokinetics causes the perpendicular terms to be an order larger than their parallel counterparts. In contrast, the lack of collisions and lack of perpendicular flows in the 1-D Vlasov Poisson system can eventually lead to non-Maxwellian filamentation in the velocity distribution function.

# Chapter 3: Numerical Methods: The Eulerian Pseudo-spectral Approach

The content in Section 3.2 of this chapter is adapted from the work published in Barbour et al. [2025], conducted in collaboration with Prof. William Dorland, Dr. Ian G. Abel, and Dr. Matt Landreman and has been reused with the permission of the surviving authors.

The electrostatic gyrokinetic and 1-D Vlasov-Poisson systems of equations derived in Chapter 2 comprise a rigorous explanation of the time evolution of the distribution function of plasmas in the limits defined in each section. Though applications of complex analysis have yielded insights about the linear instabilities supported by these systems [Catto, 1978; Hammett and Perkins, 1990], the nonlinear nature of these equations impose additional challenges. In both cases, the nonlinearity manifests as a quadratic nonlinear term involving a product between two functions of the perturbed distribution function. Although the existence of a quadratic nonlinearity in a system does not guarantee the onset of chaotic dynamics, quadratic nonlinear terms are known to be the simplest form of nonlinearity that can induce chaos [Almatroud et al., 2024]. The objectives of finding solutions to the nonlinear electrostatic gyrokinetic and Vlasov-Poisson systems can be considered as initial-value problems (IVPs) that are highly sensitive to initial conditions. This sensitivity to initial conditions necessitates numerical solutions to the PDEs.

On its face, the proposition of simulating a multi-scale, continuous system by using a numerical discretization scheme appears to be a monumental undertaking. Many questions arise. For instance, how does one represent dynamics on a continuous domain with a grid that has a finite minimum scale? The answer, evidently, is "with great care." There are a myriad of factors that must be considered when choosing a discretization scheme. A non-exhaustive list of some of these considerations includes:

- What region of physical space should the domain cover?
- What are the smallest length scales necessary to include such that the simulation faithfully represents the dynamics?
- What should be used to represent space, time, and the physical observables of the system?

We shall study the first two questions on this list in more detail in Chapter 5. In this chapter, we define and motivate our answer to the third question: pseudo-spectral methods in an Eulerian framework. We then derive the projection of our two model systems onto a suitable spectral basis.

There are three broad categories of approaches to solving PDEs in plasma physics, Lagrangian methods, Eulerian methods, and hybrid, semi-Lagrangian methods. Each approach comes with its own advantages and disadvantages. Lagrangian methods, often known as particle-in-cell (PIC) codes, track trajectories of an ensemble of markers or "macroparticles" on a spatiotemporal mesh [Birdsall and Langdon, 1991; Garbet et al., 2010]. Early gyrokinetic turbulence codes often used the Lagrangian-PIC approach, including `GTC` [Lin et al., 1998], `GT3D` [Idomura et al., 2003], `ORB5` [Jolliet et al., 2007; Lanti et al., 2020], and `GTS` [Wang et al., 2006]. With a large enough sample of particles, the statistical properties of the ensemble can be considered as representative of the collective dynamics of the system [Aydemir, 1994].

An advantage of the Lagrangian approach is that the algorithms are intuitive to implement. Macroparticle trajectories are integrated via characteristics of the fundamental equations. Fields are evaluated at points on a fixed mesh by interpolating charges and currents to the mesh points. Then, the effective fields are interpolated to the macroparticle locations, and the process repeats. The independence of the macroparticles naturally lends itself to parallelization.

A second advantage of the Lagrangian-PIC approach is that boundary interfaces between

different materials are permitted to have precise shapes. This can be particularly advantageous for vacuum vessel wall integrity simulations in the field of neutronics, as modeling collisions with these well-defined boundaries facilitates straightforward analyses of accumulated structural damage. Two commonly used codes developed for this task are the open-source `OpenMC` [Romano et al., 2015] and the industry-standard `MCNP` [Kulesza et al., 2022].

A third advantage of PIC codes is that by construction, the distribution function of the plasma, $f$, is non-negative in the entire domain. This feature is not automatically guaranteed in Eulerian simulations unless specific methods are used.

An inherent disadvantage of PIC codes is that they are highly sensitive to sampling noise. Following the Monte-Carlo interpretation of the particle distribution, the errors scale as $\sigma^2/\sqrt{N}$, where $N$ is the number of macroparticles, and $\sigma^2$ is the variance of the estimator [Garbet et al., 2010]. This implies that for a given estimator, reducing the error of a PIC simulation by a factor of two requires increasing the number of macroparticles by a factor of four. As discussed in Chapter 1, these simulations are resource-intensive, so imposing additional computational costs to counteract sampling variance is an unfortunate necessity. Garbet et al. [2010] note that typically around 100–200 macroparticles per cell are necessary to maintain a signal-to-noise ratio above 10, which in turn, is necessary for valid simulation results. In their studies of electron-temperature-gradient turbulence, Nevins et al. [2005] identified that discrete particle noise suppresses microturbulence. Furthermore, adding statistical noise to systems that are already sensitive to initial conditions is an unattractive feature of PIC simulations.

In contrast, Eulerian methods focus on static grid points in space and time and then measure changes in the statistical distribution function of the plasma at those points. This approach can be advantageous for straightforward boundary conditions in phase space, including periodic boundary

conditions or Dirichlet boundary conditions that enforce that the distribution function goes to zero at the ends of the velocity domain. This approach is often applied to gyrokinetic simulations, including `GS2` [Dorland et al., 2000; Barnes et al., 2024], `GENE` [Jenko et al., 2000; Görler et al., 2011; Germaschewski et al., 2021], `CGYRO` [Candy and Waltz, 2003; Candy et al., 2016], `GKV` [Watanabe and Sugama, 2006], and `GX` [Mandell et al., 2018, 2024]. A comparative advantage of Eulerian methods is that they do not suffer from the sampling noise introduced by PIC approaches. They also naturally accommodate spectral methods, which permits them to have more favorable convergence properties [Boyd, 2000].

A challenge associated with Eulerian methods is that they rely on numerical dissipation to promote stability in the time integration of the system of equations. This additional dissipation has the effect of smoothing out fine-scale structures, which can impede analyses of scenarios where filamentation of phase space is significant [Watanabe and Sugama, 2006]. Additionally, due to the fixed simulation grids, Eulerian simulations that must resolve sharp gradients or localized features may require more resolution than Lagrangian methods require.

Some codes, including `Gysela` [Grandgirard et al., 2006, 2016], use semi-Lagrangian approaches, which attempt to harmonize the strengths of Lagrangian and Eulerian approaches while mitigating their downsides. Specifically, they relax the CFL condition and numerical dissipation requirements of Eulerian methods while providing a structured velocity-space grid that reduces particle noise and improves the accuracy of modeling collisions in comparison to pure Lagrangian methods. The main challenge of this hybrid approach is the need to find the proper interpolation scheme that preserves particle conservation and strikes the correct balance between dissipation and overdamping [Garbet et al., 2010].

Along with finite-difference and finite-element methods, spectral methods are an approach

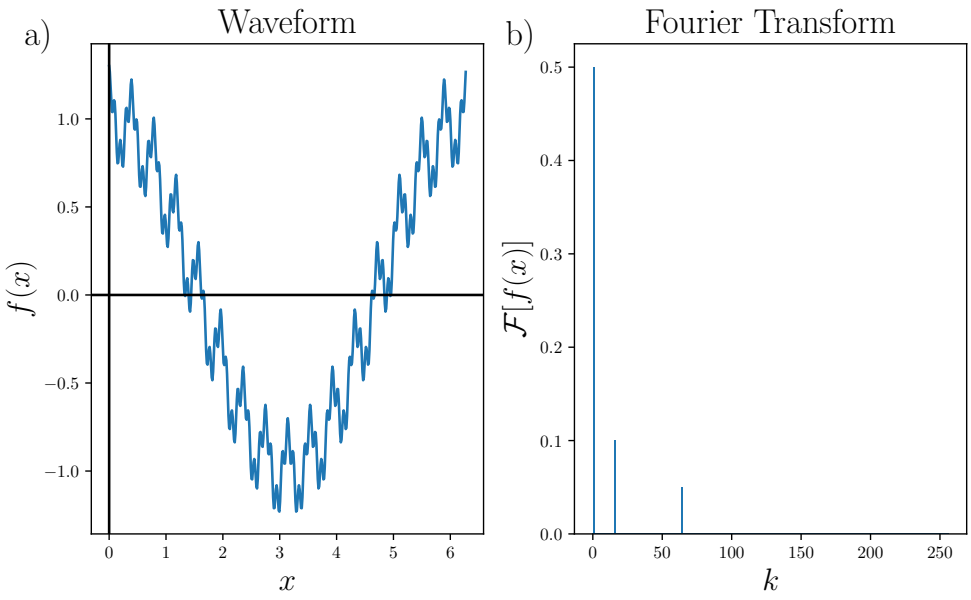



Figure 3.1: a) Cosine wave superimposed with fine-scale spatial oscillations evaluated at 512 equally spaced grid points on the domain $x \in [0, 2\pi]$: $f(x) = \cos(x) + 0.2\cos(16x) + 0.1\cos(64x)$. b) Amplitudes of the positive wavenumbers $k$ for the Fourier transform of the waveform in a).

to solving PDEs. As a broad definition, spectral methods approximate solutions to PDEs with a truncated series of "basis functions" [Mercier, 1989; Boyd, 2000]. Each basis function is global in the sense that it covers the entire domain. The spectral series then directly encodes structure information into its basis representation. A common example of a spectral basis is the Fourier series, which encodes fine-scale spatial structure as high-order trigonometric functions. Increasing the order of the basis expansion permits the basis to capture finer features. Figure 3.1 displays an oscillatory waveform composed of three superimposed Fourier harmonics. Figure 3.2 shows that filtering out the highest frequency mode smooths out the resulting inverse Fourier transform back to position space.

Finite element methods are similar to spectral methods. Their main distinction is that instead of using a series of global basis functions, finite element methods instead divide the domain into

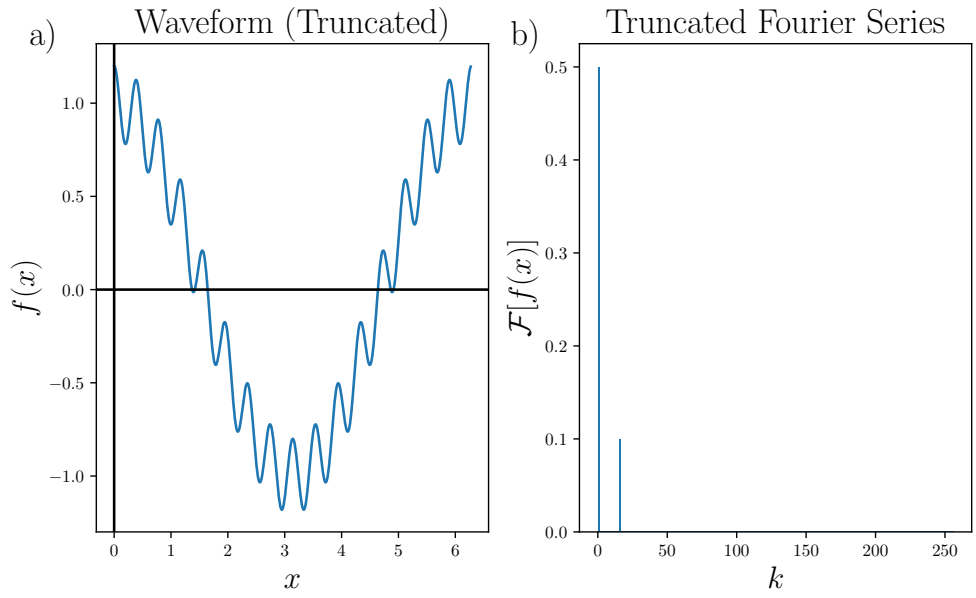



Figure 3.2: a) Inverse Fourier transform of the Fourier Series in b), which truncates the Fourier series in Figure 3.1 b) by setting all Fourier amplitudes for $|k| > 32$ to zero. The truncation removes the finest scale oscillations that resulted from the $0.1\cos(64x)$ term.

numerous sub-domains and approximate the solution using local, low-order polynomials. Finite element methods have successfully been applied to study plasma dynamics [Shiraiwa et al., 2010]. Discontinuous Galerkin methods, which are used in the 6-D Vlasov-Maxwell code `Gkyell` [Juno et al., 2018; Wang et al., 2020], can be considered as a high-order generalization of finite element and finite volume methods.

A distinct benefit of spectral methods is their accuracy and efficiency for evaluating derivatives. Pseudo-spectral methods are spectral methods where the residual is set to zero at a set of "collocation points" equal in the number to the basis functions used to represent the solution [Boyd, 2000]. This harmonizes features of grid-based methods and the global basis functions of spectral methods. The fact that the derivatives of the spectral basis functions can be calculated analytically and that the functions are global implies that the pseudo-spectral derivatives use all $N$ points in the

domain [Boyd, 2000]. In contrast, finite-difference or finite-element schemes require an $N^{th}$ order method to achieve the same accuracy as pseudo-spectral methods. Increasing the number of basis functions also decreases the corresponding grid size. This permits pseudo-spectral methods to exhibit exponential convergence, where the error scales as $\mathcal{O}[(1/N)^N]$ [Boyd, 2000]. The favorable convergence of spectral methods reduces memory requirements, and in the context of Eulerian IVP simulations, can relax the CFL condition by allowing for larger spaces between grid points at fixed accuracy.

Psuedo-spectral methods also have the benefit of naturally expressing energy cascades. According to Kolmogorov [1991], inertial ranges of fluid turbulence exhibit a common power law, where energy as a function of wavenumber scales as $k^{-5/3}$. Physically, this corresponds to large turbulent eddies breaking into successively smaller eddies, while maintaining a universal energy flux at all scales in the inertial range. A similar process occurs in the dual space of velocity for the phenomenon of phase mixing in linear Landau damping [Vlasov, 1968]. Fine scale structure emerges in velocity space without the mediating effect of collisions, and a energy cascades from large scales to small scales in velocity space. Spectral bases can directly represent fine-scale information as amplitudes of basis functions of varying orders, making them an attractive choice for our systems of interest. As we will show in Section 3.1 and Chapter 4, the ballistic free-streaming term in the kinetic equations manifests naturally as a coupling between nearest neighbors in a spectral representation for velocity. Due to its propensity to exhibit turbulent cascades and well-defined boundary conditions, the gyrokinetic system is a prime candidate for pseudo-spectral methods, and there is recent interest in developing these techniques [Jorge et al., 2017; Mandell et al., 2018, 2024; Frei et al., 2023, 2024; Hoffmann et al., 2023a,b].

In this thesis, we elect to use Eulerian pseudo-spectral methods. For our gyrokinetic turbu-

lence studies, we use GX, which solves the gyrokinetic equation on a Fourier-Laguerre-Hermite basis. We solve the Vlasov-Poisson equations on a Fourier-Hermite basis.

## 3.1 Electrostatic Gyrokinetics

In this section, we derive the projection of the gyrokinetic equation and the quasineutrality condition onto the Fourier-Laguerre-Hermite basis used in GX. This derivation follows the conventions introduced by Mandell et al. [2018]. Perpendicular lengths are normalized to the As we derived in Chapter 2, the electrostatic gyrokinetic equation uses a 5-D domain of three spatial dimensions, parallel velocity, and perpendicular velocity. The perpendicular spatial coordinates $x$ and $y$ follow periodic boundary conditions, for which a Fourier basis is suitable. The velocity coordinates decay at infinity.

We adopt their definitions of the Hermite and Laguerre basis functions, where superscripts denote expansion functions, and subscripts accompany projection functions. The Laguerre basis functions are

$$\psi^{\ell}(\mu B) \equiv (-1)^{\ell} \mathrm{e}^{-\mu B} L_{\ell}(\mu B), \quad \psi_{\ell}(\mu B) \equiv (-1)^{\ell} \mathrm{L}_{\ell}(\mu B), \tag{3.1}$$

and the Hermite basis functions are

$$\varphi^{m}(v_{\parallel}) \equiv \frac{\mathrm{e}^{-v_{\parallel}^2/2}}{\sqrt{(2\pi)^3 m!}} \mathrm{He}_m(v_{\parallel}), \quad \varphi_m(v_{\parallel}) \equiv \frac{1}{\sqrt{m!}} \mathrm{He}_m(v_{\parallel}). \tag{3.2}$$

The probabalists' Hermite ($\mathrm{He}_m$) and Laguerre ($\mathrm{L}_{\ell}$) polynomials are two families of polynomials which are orthogonal with respect to a weight function [Abramowitz and Stegun, 1972]. Instead of using the polynomials as the basis functions directly, Mandell et al. [2018] introduce nor-

malization factors into the Hermite basis functions and integration weights into both the Laguerre and Hermite basis functions so that both pairs of functions satisfy an orthonormality condition:

$$2\pi \int_{-\infty}^{\infty} \mathrm{d}v_{\parallel} \varphi_m \varphi^{m'} = \delta_{m,m'}. \tag{3.3}$$

$$\int_{0}^{\infty} \mathrm{d}\mu B \psi_{\ell} \psi^{\ell'} = \delta_{\ell,\ell'}. \tag{3.4}$$

With those definitions, we can begin with the normalized electrostatic gyrokinetic equation for a single ion species in the presence of magnetic field curvature, Equation (2.1) of Mandell et al. [2018]:

$$\frac{\partial g}{\partial t} + \left[ v_{\parallel} \widehat{\mathbf{b}} + \langle \mathbf{v}_E \rangle_{\mathbf{R}} + \frac{T_i}{Z_i} \mathbf{v}_d \right] \cdot \nabla h + \langle \mathbf{v}_E \rangle_{\mathbf{R}} \cdot \nabla F_M - \mu (\widehat{\mathbf{b}} \cdot \nabla B) \frac{\partial h}{\partial v_{\parallel}} = C[h], \tag{3.5}$$

where we introduce the $g(\mathbf{R}, v_{\parallel}, \mu, t)$ as the gyroaverage of $\delta f_1$ from Section 2.2.4,

$$\langle F \rangle_{\mathbf{R}} = F_M(\mathbf{R}) + g, \tag{3.6}$$

$$g = h - \frac{Z_i}{T_i} \langle \Phi \rangle_{\mathbf{R}} F_M. \tag{3.7}$$

In (3.5) and in this section, we define $\nabla$ to be $(\partial / \partial \mathbf{R})$. Perpendicular velocity has been transformed into the adiabatic invariant $\mu$, defined by

$$\mu = \frac{v_{\perp}^2}{2B}, \tag{3.8}$$

where $B \equiv |\mathbf{B}|$ as in Chapter 2. The gyrophase-independent portion of the perturbed distribution

function can be expanded in terms of the Laguerre-Hermite basis functions:

$$h(\mathbf{R}, v_\parallel, \mu) = \sum_{\ell=0}^{\infty} \sum_{m=0}^{\infty} \psi^\ell(\mu B)\varphi^m(v_\parallel) H_{\ell,m}(\mathbf{R}). \tag{3.9}$$

We shall project (3.5) onto the Laguerre-Hermite basis term-by-term, defining the Laguerre-Hermite projection operator of a function $g$ as

$$\mathcal{H}_m \mathcal{L}_\ell[g] \equiv 2\pi \int_{-\infty}^{\infty} \int_0^{\infty} \mathrm{d}v_\parallel \mathrm{d}\mu B \varphi_m \psi_\ell g. \tag{3.10}$$

The explicit time derivative term is the simplest term of (3.5) to project, using the orthonormality relation (3.77):

$$\mathcal{H}_m \mathcal{L}_\ell \frac{\partial g}{\partial t} = 2\pi \frac{\partial}{\partial t} \int_{-\infty}^{\infty} \int_0^{\infty} \mathrm{d}v_\parallel \mathrm{d}\mu B \varphi_m \psi_\ell \sum_{\ell',m'} \psi^{\ell'} \varphi^m G_{\ell',m'} = \frac{\partial G_{\ell,m}}{\partial t}. \tag{3.11}$$

The next term is the ballistic free-streaming term, $v_\parallel \widehat{\mathbf{b}} \cdot \nabla h$. To complete the projection, we will need to derive several identities for individual terms in the expansion. The gradient of $h$ with respect to gyrocenter position at fixed $\mu$ is

$$\begin{aligned} \nabla h &= \nabla \left( \sum_{\ell,m} \psi^\ell \varphi^m H_{\ell,m} \right) \\ &= \sum_{\ell,m} \varphi^m \left( H_{\ell,m} \nabla \psi^\ell + \psi^\ell \nabla H_{\ell,m} \right). \end{aligned} \tag{3.12}$$

Note that the gradient of each Laguerre basis function at fixed $\mu$ is nonzero due to the magnetic

field magnitude in its argument:

$$\begin{aligned}\nabla\psi^{\ell} &= \frac{\partial\psi^{\ell}}{\partial(\mu B)}\nabla(\mu B) \\ &= \mu\frac{\partial\psi^{\ell}}{\partial(\mu B)}\nabla B.\end{aligned} \tag{3.13}$$

To complete the expansion of $\nabla\psi^{\ell}$, we must compute the derivative of the Laguerre expansion function $\psi^{\ell}$ with respect to its argument, $\mu B$:

$$\begin{aligned}\frac{\partial\psi^{\ell}}{\partial(\mu B)} &= \frac{\partial}{\partial(\mu B)}\left[(-1)^{\ell}\mathrm{e}^{-\mu B}\mathrm{L}_{\ell}\right] \\ &= (-1)^{\ell}\left[-\mathrm{e}^{-\mu B}\mathrm{L}_{\ell} + \mathrm{e}^{-\mu B}\frac{\partial}{\partial(\mu B)}\mathrm{L}_{\ell}\right].\end{aligned} \tag{3.14}$$

We can use a derivative identity and recurrence relation for the Laguerre polynomials to express the derivative of the Laguerre polynomial in (3.14) in terms of Laguerre polynomials of different orders. According to Arfken et al. [2013], the derivative identity for Laguerre polynomials is

$$x\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{L}_{\ell}(x) = \ell\mathrm{L}_{\ell}(x) - \ell\mathrm{L}_{\ell-1}(x). \tag{3.15}$$

Laguerre polynomials also follow a recurrence relation,

$$(\ell+1)\mathrm{L}_{\ell+1}(x) = (2\ell+1-x)\mathrm{L}_{\ell}(x) - \ell\mathrm{L}_{\ell-1}(x), \tag{3.16}$$

which we can rearrange to obtain an expression for the final term in the derivative identity:

$$-\ell\mathrm{L}_{\ell-1}(x) = (\ell+1)\mathrm{L}_{\ell+1}(x) - (2\ell+1-x)\mathrm{L}_{\ell}(x). \tag{3.17}$$

Substituting (3.17) into the derivative identity (3.15) yields

$$\begin{aligned} x\frac{\mathrm{d}}{\mathrm{d}x}L_\ell &= \ell \mathrm{L}_\ell + (\ell+1)\mathrm{L}_{\ell+1} - (2\ell+1-x)\mathrm{L}_\ell \\ &= x\mathrm{L}_\ell + (\ell+1)[\mathrm{L}_\ell - \mathrm{L}_{\ell+1}]. \end{aligned} \tag{3.18}$$

Finally, we can use this result to evaluate the original derivative of each Laguerre expansion function (3.14) as a sum of two expansion functions, where the argument of the Laguerre polynomials is again $\mu B$:

$$\begin{aligned} \frac{\partial \psi^\ell}{\partial(\mu B)} &= (-1)^\ell \mathrm{e}^{-\mu B}\frac{(\ell+1)}{\mu B}\left[\mathrm{L}_\ell - \mathrm{L}_{\ell+1}\right] \\ &= -\frac{(\ell+1)}{\mu B}[\psi^\ell + \psi^{\ell+1}], \end{aligned} \tag{3.19}$$

where in the final line, we used the definition of the Laguerre expansion functions (3.1). With this intermediate identity derived, we can return to (3.13) and compute the gradient of the Laguerre expansion functions at fixed $\mu$.

$$\begin{aligned} \nabla\psi^\ell &= \mu\frac{\partial\psi^\ell}{\partial(\mu B)}\nabla B \\ &= \mu\left(-\frac{(\ell+1)}{\mu B}[\psi^\ell + \psi^{\ell+1}]\right)\nabla B \\ &= -(\ell+1)[\psi^\ell + \psi^{\ell+1}]\nabla \ln B. \end{aligned} \tag{3.20}$$

Finally, we combine (3.20) and (3.21) to form a final expression for $\nabla h$ at fixed $\mu$:

$$\nabla h = \sum_{\ell,m}\varphi^m(\psi^\ell \nabla H_{\ell,m} - (\ell+1)[\psi^\ell + \psi^{\ell+1}]H_{\ell,m}\nabla \ln B). \tag{3.21}$$

Thus, the parallel derivative in (3.5) becomes

$$\widehat{\mathbf{b}} \cdot \nabla h = \sum_{\ell,m} \varphi^m (\psi^\ell (\widehat{\mathbf{b}} \cdot \nabla) H_{\ell,m} - (\ell+1)[\psi^\ell + \psi^{\ell+1}] H_{\ell,m} (\widehat{\mathbf{b}} \cdot \nabla) \ln B). \tag{3.22}$$

Next, we wish to remove the explicit dependence on $v_\parallel$ from the free streaming term. To accomplish this task, we will use the recurrence relation for Hermite polynomials [Abramowitz and Stegun, 1972]:

$$x \mathrm{He}_m(x) = \mathrm{He}_{m+1}(x) + m \mathrm{He}_{m-1}(x). \tag{3.23}$$

When (3.23) is rewritten in terms of our Hermite expansion functions (3.2), we find the following recurrence relation:

$$v_\parallel \varphi^m(v_\parallel) = \sqrt{m+1} \varphi^{m+1}(v_\parallel) + \sqrt{m} \varphi^{m-1}(v_\parallel). \tag{3.24}$$

As promised in the introduction to this chapter, combining the Hermite recurrence relation (3.24) with our expression for $\nabla h$ (3.12) reveals that the parallel streaming term couples nearest neighbors in Hermite space:

$$\begin{aligned} v_\parallel \widehat{\mathbf{b}} \cdot \nabla h &= \sum_{\ell,m} v_\parallel \varphi^m \{\psi^\ell \nabla_\parallel H_{\ell,m} - (\ell+1)[\psi^\ell + \psi^{\ell+1}] H_{\ell,m} \nabla_\parallel \ln B\} \\ &= \sum_{\ell,m} (\sqrt{m+1} \varphi^{m+1} + \sqrt{m} \varphi^{m-1}) \{\psi^\ell \nabla_\parallel H_{\ell,m} - (\ell+1)[\psi^\ell + \psi^{\ell+1}] H_{\ell,m} \nabla_\parallel \ln B\}, \end{aligned} \tag{3.25}$$

where we have introduced the notation $\nabla_\parallel$ for the spatial derivative parallel to the magnetic field:

$$\nabla_\parallel \equiv \widehat{\mathbf{b}} \cdot \nabla. \tag{3.26}$$

The Laguerre-Hermite projection of the free streaming term is

$$\begin{aligned}\mathcal{H}_m\mathcal{L}_\ell[v_\parallel\nabla_\parallel h] = &(\sqrt{m+1}\varphi^{m+1} + \sqrt{m}\varphi^{m-1})\nabla_\parallel H_{\ell,m} \\ &- [\ell\sqrt{m+1}H_{\ell-1,m-1} + \ell\sqrt{m+1}H_{\ell-1,m+1} \\ &+ (\ell+1)\sqrt{m+1}H_{\ell,m+1} + (\ell+1)\sqrt{m}H_{\ell,m-1}]\nabla_\parallel \ln B.\end{aligned} \tag{3.27}$$

Although parallel variations share a common reference scale, the small-scale distribution function varies on the order of the ion Larmor radius $\rho_i$, while gradients in the magnetic field have the macroscopic reference length scale, $a$. Therefore, in our formal expansion where $\rho_i \ll a$, the perpendicular portion of the second term of $\nabla h$ (3.12) is an order smaller than the first term. We shall neglect it:

$$\widehat{\mathbf{b}} \times \nabla h = \widehat{\mathbf{b}} \times \sum_{\ell,m} \varphi^m \psi^\ell \nabla H_{\ell,m} + \mathcal{O}\left(\frac{\epsilon b_s}{\rho_i}\right). \tag{3.28}$$

The magnetic drifts $\mathbf{v}_d$ comprise the next terms to project onto the basis,

$$\mathbf{v}_d \equiv (v_\parallel^2 + \mu B)\frac{\widehat{\mathbf{b}} \times \nabla B}{B^2}. \tag{3.29}$$

We define the curvature drift operator as

$$\mathrm{i}\omega_d^\kappa \equiv \frac{1}{B}\widehat{\mathbf{b}} \times (\widehat{\mathbf{b}} \cdot \nabla\widehat{\mathbf{b}}) \cdot \nabla \tag{3.30}$$

such that the advection of $h$ by the curvature drift becomes $v_\parallel^2 \mathrm{i}\omega_d^\kappa h$. This curvature drift term

requires $v_\parallel^2 h$, which we expand with two applications of (3.24):

$$\begin{aligned} v_\parallel^2 h &= \sum_m v_\parallel^2 \varphi^m H_{\ell,m} \\ &= \sum_m v_\parallel \left( \sqrt{m+1}\varphi^{m+1} + \sqrt{m}\varphi^{m-1} \right) H_{\ell,m} \\ &= \sum_m \left[ \sqrt{(m+1)(m+2)}\varphi^{m+2} + (2m+1)\varphi^m + \sqrt{m(m-1)}\varphi^{m-2} \right] H_{\ell,m}. \end{aligned} \tag{3.31}$$

The projection of the curvature drift onto the Laguerre-Hermite basis is

$$\mathcal{H}_m \mathcal{L}_\ell \mathrm{i}\omega_D^\kappa v_\parallel^2 h = \mathrm{i}\omega_D^\kappa (\sqrt{m(m-1)} H_{\ell,m-2} + (2m+1) H_{\ell,m} + \sqrt{(m+1)(m+2)} H_{\ell,m+2}). \tag{3.32}$$

Next, we introduce the magnetic field gradient drift operator

$$\mathrm{i}\omega_D^{\nabla B} = \frac{1}{B^2} (\widehat{\mathbf{b}} \times \nabla B) \cdot \nabla, \tag{3.33}$$

such that the $\nabla B$ drift can be expressed as $\mu B \mathrm{i}\omega_D^{\nabla B} h$. We can use the recurrence relation for Laguerre polynomials (3.17) to derive the corresponding recurrence relation for the Laguerre expansion functions:

$$\begin{aligned} (\ell+1)\mathrm{L}_{\ell+1}(x) &= (2\ell+1-x)\mathrm{L}_\ell(x) - \ell \mathrm{L}_{\ell-1}(x) \\ \mu B L_\ell &= (2\ell+1)L_\ell - (\ell+1)L_{\ell+1} - \ell L_{\ell-1} \\ \mu B \mathrm{e}^{\mu B}(-1)^\ell \psi^\ell &= \mathrm{e}^{\mu B}\left[ (2\ell+1)(-1)^\ell \psi^\ell - (\ell+1)(-1)^{\ell+1}\psi^{\ell+1} - \ell(-1)^{\ell-1}\psi^{\ell-1} \right] \\ \mu B \psi^\ell &= (2\ell+1)\psi^\ell + (\ell+1)\psi^{\ell+1} + \ell\psi^{\ell-1}. \end{aligned} \tag{3.34}$$

Then, we substitute this result into the $\nabla B$ drift term and project it onto the Laguerre-Hermite

basis:

$$\begin{aligned}\mathcal{H}_m\mathcal{L}_\ell[\mathrm{i}\omega_d^{\nabla B}\mu Bh] &= \mathrm{i}\omega_d^{\nabla B}\int_0^\infty \mathrm{d}(\mu B)\psi_\ell \sum_{\ell'}\mu B\psi^{\ell'}H_{\ell',m} \\ &= \mathrm{i}\omega_d^{\nabla B}\int_0^\infty \mathrm{d}(\mu B)\psi_\ell \sum_{\ell'}[(2\ell'+1)\psi^{\ell'}+(\ell'+1)\psi^{\ell'+1}+\ell\psi^{\ell-1}]H_{\ell',m} \\ &= \mathrm{i}\omega_d^{\nabla B}[(2\ell+1)H_{\ell,m}+(\ell+1)H_{\ell+1,m}+\ell H_{\ell-1,m}].\end{aligned} \tag{3.35}$$

Next we consider the bounce term,

$$\mu(\widehat{\mathbf{b}}\cdot\nabla B)\frac{\partial h}{\partial v_\parallel} = \nabla_\parallel(\ln B)\mu B\frac{\partial h}{\partial v_\parallel}. \tag{3.36}$$

To project the parallel velocity derivative onto the Hermite basis, we will need an expression for the derivative of the Hermite polynomials. These derivatives follow the sequence [Arfken et al., 2013]

$$\frac{\mathrm{d}}{\mathrm{d}v_\parallel}\mathrm{He}_m(v_\parallel) = m\mathrm{He}_{m-1}(v_\parallel). \tag{3.37}$$

Therefore, we can derive the corresponding sequence for the Hermite expansion functions:

$$\begin{aligned}\frac{\partial}{\partial v_\parallel}\varphi^m &= \frac{\partial}{\partial v_\parallel}\left[\frac{\mathrm{e}^{-v_\parallel^2/2}}{\sqrt{(2\pi)^3 m!}}\mathrm{He}_m\right] \\ &= \frac{\mathrm{e}^{-v_\parallel^2/2}}{\sqrt{(2\pi)^3 m!}}\left[-v_\parallel\mathrm{He}_m+\frac{\partial}{\partial v_\parallel}\mathrm{He}_m\right] \\ &= \frac{\mathrm{e}^{-v_\parallel^2/2}}{\sqrt{(2\pi)^3 m!}}\left[-v_\parallel\mathrm{He}_m+m\mathrm{He}_{m-1}\right] \\ &= \frac{\mathrm{e}^{-v_\parallel^2/2}}{\sqrt{(2\pi)^3 m!}}\left[-\mathrm{He}_{m+1}\right] \\ &= -\sqrt{m+1}\varphi^{m+1}.\end{aligned} \tag{3.38}$$

We can then use the recurrence relation (3.23) to expand the bounce term in terms of the Laguerre-Hermite expansion functions:

$$\begin{aligned}\nabla_\parallel(\ln B)\mu B\frac{\partial h}{\partial v_\parallel} &= \nabla_\parallel(\ln B)\sum_{\ell,m}\mu B\psi^\ell\frac{\partial}{\partial v_\parallel}\varphi^m H_{\ell,m} \\ &= \nabla_\parallel(\ln B)\sum_{\ell,m}[(2\ell+1)\psi^\ell + (\ell+1)\psi^{\ell+1} + \ell\psi^{\ell-1}](-\sqrt{m+1})\varphi^{m+1}H_{\ell,m}.\end{aligned} \tag{3.39}$$

Projecting this term onto the Laguerre-Hermite basis using the orthonormality relations (3.3 - 3.4) yields another local interaction in velocity space:

$$\nabla_\parallel(\ln B)\mathcal{H}_m\mathcal{L}_\ell\left[\mu B\frac{\partial h}{\partial v_\parallel}\right] = -\nabla_\parallel(\ln B)\sqrt{m}[(2\ell+1)H_{\ell,m-1} + \ell H_{\ell-1,m-1} + (\ell+1)H_{\ell+1,m-1}] \tag{3.40}$$

Next, we will project the drive term, $\langle \mathbf{v}_E\rangle_{\mathbf{R}}\cdot\nabla F_M$, onto the basis. Therefore, we will need an expression for the gradient of the background Maxwellian,

$$F_M = \frac{n}{(2\pi v_t^2)^{3/2}}\mathrm{e}^{-v_\parallel^2/2-\mu B}, \tag{3.41}$$

where all quantities have been normalized. Using the chain rule,

$$\begin{aligned}\nabla F_M &= \nabla n\frac{\partial F_M}{\partial n} + \nabla T\frac{\partial v_t}{\partial T}\left(\frac{\partial F_M}{\partial v_t} + \frac{\partial v_\parallel}{\partial v_t}\frac{\partial F_M}{\partial v_\parallel} + \frac{\partial(\mu B)}{\partial v_t}\frac{\partial F_M}{\partial(\mu B)}\right) \\ &= \frac{\nabla n}{n}F_M + \frac{\nabla T}{2T}v_t\left[-\frac{3}{v_t}F_M + \frac{v_\parallel^2}{v_t}F_M + \frac{2\mu B}{v_t}F_M\right] \\ &= \left(\frac{a}{L_n} + \frac{a}{L_T}\left[\frac{v_\parallel^2}{2} + \mu B - \frac{3}{2}\right]\right)F_M\nabla x,\end{aligned} \tag{3.42}$$

where we have introduced the normalized density and temperature gradients, $a/L_n$ and $a/L_T$.

The final line in (3.42) reflects the toroidal geometry of our magnetic fusion simulations, as we choose to use field-aligned spatial coordinates in accordance with the formalism developed by Beer et al. [1995]. In a toroidal flux tube formulation of gyrokinetics, the equilibrium temperature and density are flux functions, which constant on a surface of constant magnetic flux. We use the magnetic coordinates defined by the Clebsch representation for the magnetic field [D'haeseleer et al., 1991],

$$\mathbf{B} = \nabla\alpha \times \nabla\psi, \tag{3.43}$$

which is a consequence of the divergence-free condition for magnetic fields. The radially-local flux tube in the formalism of Beer et al. [1995] is defined by a non-orthogonal coordinate system of flux surface label $x(\psi)$, field line label $y(\alpha)$, and a field-line-following coordinate $z(\theta)$ that uses the poloidal angle $\theta$ to measure distance along a magnetic field line. Gradients of equilibrium quantities can then be expressed as a product between the corresponding gradient scale lengths, which are constant along the flux surface, and the radial gradient $\nabla x$, which varies along the field line.

We can use the definition of the Hermite polynomials to express the $v_\parallel^2 F_M$ term in $\nabla F_M$ (3.42) as a sum of Hermite expansion functions, which eliminates explicit dependence on $v_\parallel$:

$$\begin{aligned}\frac{v_\parallel^2}{2}F_M &= \left(\frac{\mathrm{He}_2(v_\parallel)+\mathrm{He}_0(v_\parallel)}{2}\right)\frac{n}{(2\pi v_t^2)^{3/2}}\mathrm{e}^{-v_\parallel^2/2-\mu B} \\ &= \frac{n}{v_t^3}\left(\frac{1}{\sqrt{2}}\varphi^2+\frac{1}{2}\varphi^0\right)\mathrm{e}^{-\mu B}.\end{aligned} \tag{3.44}$$

Similarly, we can expand the $\mu B F_M$ term in (3.42):

$$\begin{aligned}\mu B F_M &= [L_0(\mu B) - L_1(\mu B)]\frac{n}{(2\pi v_t^2)^{3/2}}\mathrm{e}^{-v_\parallel^2/2-\mu B} \\ &= [\psi^1 + \psi^0]\frac{n}{v_t^3}.\end{aligned} \tag{3.45}$$

$$\langle \mathbf{v}_E \rangle_{\mathbf{R}} = \widehat{\mathbf{b}} \times \nabla \langle \Phi \rangle_{\mathbf{R}}. \tag{3.46}$$

Similar to the drift terms, we define the geometric coefficient of the drive term as

$$\mathrm{i}\widetilde{\omega}_* \equiv \mathrm{i}k_y. \tag{3.47}$$

We can then write the drive term as

$$-\langle \mathbf{v}_E \rangle_{\mathbf{R}} \cdot \nabla F_M = \mathrm{i}\widetilde{\omega}_* \langle \Phi \rangle_{\mathbf{R}} \frac{\partial F_M}{\partial x}. \tag{3.48}$$

Note that the gyroaverage operation introduces perpendicular velocity dependence into the drive term. We must be careful to account for this when performing the projection onto the Laguerre-Hermite basis. Although we shall soon prove that the gyroaverage operation has a simple representation in Fourier space, its Laguerre representation is dense. Following Mandell et al. [2024], we define the Laguerre projection of the gyroaverage operator as

$$\mathcal{J}_\ell \equiv \int_0^\infty \mathrm{d}\mu B \psi_\ell J_0(\sqrt{2\mu B b_i})\mathrm{e}^{-\mu B} = \int_0^\infty \mathrm{d}\mu B \psi^\ell J_0(\sqrt{2\mu B b_i}), \tag{3.49}$$

where the argument of the Bessel function of the first kind includes the dimensionless quantity $b_i$,

$$b_i \equiv k_\perp^2 \rho_i^2, \tag{3.50}$$

and the thermal ion gyroradius is

$$\rho_i \equiv \frac{m_i v_t}{Z_i B}. \tag{3.51}$$

The Laguerre-Hermite projection of the drive term is therefore

$$\begin{aligned}
\mathcal{D}_{\ell,m} &\equiv \mathcal{H}_m \mathcal{L}_\ell [-\langle \mathbf{v}_E \rangle_{\mathbf{R}} \cdot \nabla F_M] \\
&= \mathrm{i}\widetilde{\omega}_* 2\pi \int_0^\infty \int_{-\infty}^\infty \mathrm{d}v_\parallel \mathrm{d}\mu B \psi_\ell J_0 \Phi \left( \frac{a}{L_n} + \frac{a}{L_T} \left[ \frac{v_\parallel^2}{2} + \mu B - \frac{3}{2} \right] \right) F_M \\
&= \mathrm{i}\widetilde{\omega}_* \frac{n}{v_t^3} \left\{ \frac{a}{L_n} \delta_{m,0} \mathcal{J}_\ell + \frac{a}{L_T} \left[ \mathcal{J}_\ell \left( \frac{1}{\sqrt{2}} \delta_{m,2} - \delta_{m,0} \right) \right. \right. \\
&\qquad \left. \left. + \int_0^\infty \mathrm{d}\mu B J_0 \mu B \psi^\ell \delta_{m,0} \right] \right\} \Phi \\
&= \mathrm{i}\widetilde{\omega}_* \frac{n}{v_t^3} \left\{ \frac{a}{L_n} \delta_{m,0} \mathcal{J}_\ell + \frac{a}{L_T} \left[ \mathcal{J}_\ell \left( \frac{1}{\sqrt{2}} \delta_{m,2} - \delta_{m,0} \right) \right. \right. \\
&\qquad \left. \left. + \ \delta_{m,0} \int_0^\infty \mathrm{d}\mu B J_0 (2\ell+1)\psi^\ell + (\ell+1)\psi^{\ell+1} + \ell \psi^{\ell-1} \right] \right\} \Phi \\
&= \mathrm{i}\widetilde{\omega}_* \frac{n}{v_t^3} \left\{ \frac{a}{L_n} \delta_{m,0} \mathcal{J}_\ell + \frac{a}{L_T} \left[ \frac{1}{\sqrt{2}} \delta_{m,2} \mathcal{J}_\ell \right. \right. \\
&\qquad \left. \left. + \delta_{m,0} \left[ (\ell+1)\mathcal{J}_{\ell+1} 2\ell \mathcal{J}_\ell + \ell \mathcal{J}_{\ell-1} \right] \right] \right\} \Phi,
\end{aligned} \tag{3.52}$$

where we have used (3.34) to expand $\mu B \psi^\ell$ to eliminate explicit dependence on $\mu$.

The next term for us to consider is the collision term, $\mathcal{C}[h]$. There are a plethora of collision operators that have been developed to model collisions between particle species, each with varying levels of physics fidelity. The most fundamental classical description of collisions is the Boltzmann

collision operator [Boltzmann, 1872]. In the context of core turbulence in magnetic confinement fusion plasmas, the high-ionization and weak coupling assumptions lead to predominantly small-angle, glancing collisions, as opposed to the direct, hard-body collisions modeled by the Boltzmann collision operator. These collisions can be modeled by a Fokker-Planck collision operator. The full Fokker-Planck collision operator in the Landau form is given by [Landau, 1936]

$$\mathcal{C}_{FP}[f_s, f_{s'}] = -\frac{\gamma_{s,s'}}{m_s}\frac{\partial}{\partial \mathbf{v}}\left\{\int \mathrm{d}^3\mathbf{v}' \frac{\partial^2 |\mathbf{v}-\mathbf{v}'|}{\partial \mathbf{v}\partial \mathbf{v}}\left[\frac{f_{s'}(\mathbf{v}')}{m_s}\frac{\partial f_s}{\partial \mathbf{v}} - \frac{f_s(\mathbf{v})}{m_{s'}}\frac{\partial f_{s'}}{\partial \mathbf{v}}\right]\right\}, \tag{3.53}$$

where

$$\gamma_{s,s'} = 2\pi Z_s^2 Z_{s'}^2 e^4 \ln \Lambda_{s,s'}, \tag{3.54}$$

and $\ln \Lambda_{s,s'}$ is the Coulomb logarithm [Rosenbluth et al., 1957; Helander and Sigmar, 2002; Hardman et al., 2025]. This collision operator is expensive to calculate, as it is an integro-differential operator that has coefficients that depend on the distribution function. For faster evaluation of the collision operator, we would prefer to use a model that uses local diffusion. Going one step further, a collision operator whose eigenfunctions are elements of our spectral basis would allow collisions to take the form of multiplication by its eigenvalues. The GX code uses a gyroaveraged generalization of the Dougherty collision operator [Dougherty, 1964]:

$$C[\delta f] = \nu \frac{\partial}{\partial \mathbf{v}} \cdot \left[\frac{T_0}{M}\frac{\partial \delta f}{\partial \mathbf{v}} + \frac{\delta T}{M}\frac{\partial F_M}{\partial \mathbf{v}} + \mathbf{v}\delta f - \delta \mathbf{u} F_M\right], \tag{3.55}$$

where $\delta \mathbf{u}$ is the flow perturbation [Mandell et al., 2018]. The projection onto the Laguerre-Hermite basis for $m > 2$ is

$$\mathcal{C}_{\ell,m} = -\nu(b + 2\ell + m)H_{\ell,m}. \tag{3.56}$$

The unknowns $\delta T$ and $\delta n$ are low-order moments of $\delta f$ that ensure energy and momentum conservation.

The final term that remains is the nonlinear term, $\langle \mathbf{v}_E \rangle_{\mathbf{R}} \cdot \nabla h$. To calculate this, we will need the gyroaveraged $\mathbf{E} \times \mathbf{B}$ velocity:

$$\begin{aligned} \langle \mathbf{v}_E \rangle_{\mathbf{R}} &= \hat{\mathbf{b}} \times \langle \nabla \Phi \rangle_{\mathbf{R}} \\ &= \left\langle -\frac{\partial \Phi}{\partial y} \hat{\mathbf{x}} + \frac{\partial \Phi}{\partial x} \hat{\mathbf{y}} \right\rangle_{\mathbf{R}}. \end{aligned} \tag{3.57}$$

We can now introduce the Fourier transform of the electrostatic potential:

$$\begin{aligned} \mathcal{F}_{k_\perp}[\Phi(\mathbf{r})] &= \int \int \mathrm{d}x \mathrm{d}y \mathrm{e}^{-\mathrm{i}\mathbf{k}_\perp \cdot \mathbf{R}} \Phi(\mathbf{r}) \\ &\equiv \Phi_k(z). \end{aligned} \tag{3.58}$$

Its corresponding inverse Fourier transform is

$$\begin{aligned} \mathcal{F}_{k\perp}^{-1}[\Phi_k(z)] &= \sum_{k_\perp} \mathrm{e}^{\mathrm{i}\mathbf{k}_\perp \cdot \mathbf{R}} \Phi_k(z) \\ &\equiv \Phi(\mathbf{r}). \end{aligned} \tag{3.59}$$

As mentioned before, in the Fourier representation, the gyroaverage operation can manifest as multiplication by a Bessel function. We begin to prove this by first expressing the electrostatic potential in the Fourier basis. Note that the potential is a function of the particle position $\mathbf{r}$, while we choose the guiding center position as the spatial variable for the distribution function. Therefore, we introduce the intermediate step of converting from $\mathbf{r}$ to $\mathbf{R}+\boldsymbol{\rho}$, where $\boldsymbol{\rho}$ is a vector pointing

from the guiding center to the particle position.

$$\begin{aligned}\langle\Phi(\mathbf{r})\rangle_{\mathbf{R}} &\equiv \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta\Phi(\mathbf{r}) \\ &= \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta\Phi(\mathbf{R}+\mathbf{r}) \\ &= \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta\sum_{\mathbf{k}_\perp}\Phi_k(z)\mathrm{e}^{\mathrm{i}\mathbf{k}_\perp\cdot(\mathbf{R}+\mathbf{r})}.\end{aligned} \tag{3.60}$$

Next, we define the vector $\boldsymbol{\rho}$ in terms of the gyroangle $\theta$:

$$\boldsymbol{\rho} = \rho_s\cos\theta\hat{x}' + \rho_s\sin\theta\hat{y}', \tag{3.61}$$

where $\rho_s$ is the Larmor radius of the particle species. Note that we have freedom to choose the origin of the coordinate systems. We can therefore align the $x'$ coordinate of $\mathbf{r}$ with $\mathbf{k}_\perp$, which yields

$$\mathbf{k}_\perp\cdot\mathbf{r} = k_\perp\rho_s\cos\theta. \tag{3.62}$$

We can then substitute (3.62) into (3.60), which leads to

$$\langle\Phi(\mathbf{r})\rangle_{\mathbf{R}} = \frac{1}{2\pi}\int_0^{2\pi}\mathrm{d}\theta\sum_{\mathbf{k}_\perp}\Phi_k(z)\mathrm{e}^{\mathrm{i}(\mathbf{k}_\perp\cdot\mathbf{R})}\mathrm{e}^{\mathrm{i}(k_\perp\rho_s\cos\theta)}. \tag{3.63}$$

Then, we can reorder the operations such that the integral over $\theta$ occurs first:

$$\langle\Phi(\mathbf{r})\rangle_{\mathbf{R}} = \sum_{\mathbf{k}_\perp}\Phi_k(z)\mathrm{e}^{\mathrm{i}\mathbf{k}_\perp\cdot\mathbf{R}}\frac{1}{2\pi}\int_0^{2\pi}\mathrm{d}\theta\mathrm{e}^{\mathrm{i}k_\perp\rho_s\cos\theta}. \tag{3.64}$$

Finally, we observe that the remaining integral follows the form of a Bessel function of the first

kind, $J_0$ Arfken et al. [2013]:

$$\langle \Phi(\mathbf{r}) \rangle_{\mathbf{R}} = \sum_{\mathbf{k}_\perp} \Phi_k(z) \mathrm{e}^{\mathrm{i}\mathbf{k}_\perp \cdot \mathbf{R}} J_0(k_\perp \rho_s). \tag{3.65}$$

Taking the Fourier transform of this expression yields

$$\mathcal{F}_{k_\perp}[\langle \Phi(\mathbf{r}) \rangle_{\mathbf{R}}] = \Phi_k J_0(k_\perp \rho_s). \tag{3.66}$$

Combining this expression for the gyroaverage of the potential in the Fourier basis with (3.57) results in the functional form of the $\mathbf{E} \times \mathbf{B}$ velocity in the Fourier basis:

$$\mathcal{F}_{k_\perp}[\langle \mathbf{v}_E \rangle_{\mathbf{R}}] = -\mathrm{i}k_y J_0(k_\perp \rho_s) \Phi_k \hat{\mathbf{x}} + \mathrm{i}k_x J_0(k_\perp \rho_s) \Phi_k \hat{\mathbf{y}}. \tag{3.67}$$

We now desire the Hermite-Laguerre projection of this equation to match the formalism used in GX. Mandell et al. [2018] define the species-specific gyroaverage operator as

$$J_{0s} = J_0(\sqrt{2\mu B b_s}), \tag{3.68}$$

where $b_s = k_\perp^2 \rho_s^2$, and $\rho_s$ is the thermal gyroradius of species $s$. Using the Laguerre projection of the Bessel function (3.49), we find that the Fourier-Hermite-Laguerre projection of $\langle \mathbf{v}_E \rangle_{\mathbf{R}}$ becomes

$$\mathcal{L}_\ell \mathcal{H}_m \mathcal{F}_{k_\perp}[\langle \mathbf{v}_E \rangle_{\mathbf{R}}] = -\mathrm{i}k_y \mathcal{J}_\ell^s \Phi_k \delta_{m0} \hat{\mathbf{x}} + \mathrm{i}k_x \mathcal{J}_\ell^s \Phi_k \delta_{m0} \hat{\mathbf{y}}. \tag{3.69}$$

The projection of the gradient of the perturbed distribution function is straightforward to evaluate.

Because the $\mathbf{E}\times\mathbf{B}$ velocity is orthogonal to the magnetic field, we can calculate $\nabla_\perp h_s$, recognizing that the nonlinear term does not contain $\nabla_\parallel$:

$$\mathcal{L}_\ell \mathcal{H}_m \mathcal{F}_{k_\perp}[\nabla_\perp h_s] = \mathrm{i}k_x H_{\ell m}\hat{\mathbf{x}} + \mathrm{i}k_y H_{\ell m}\hat{\mathbf{y}} \tag{3.70}$$

We evaluate the nonlinear term using a pseudo-spectral approach. Note that a fully spectral representation would introduce coupling in both the Laguerre and Fourier indices because the argument of the gyroaverage operator is dense in $k_\perp$ and $\ell$. Therefore, products will be evaluated in configuration space via the inverse Fourier and Laguerre transforms. We can write the electrostatic nonlinear term as

$$\begin{aligned}\mathcal{N}^{GK}_{\ell,m} \equiv \mathcal{L}_\ell \mathcal{H}_m \mathcal{F}_{k_\perp}[\langle \mathbf{v}_E\rangle_{\mathbf{R}}\cdot\nabla h_s] = \mathcal{L}_\ell \mathcal{F}_{k_\perp}\Bigg[ &- \mathcal{F}^{-1}_{k_\perp}[\mathrm{i}k_y J_{0s}\Phi_k\delta_{m0}]\mathcal{F}^{-1}_{k_\perp}\left[\mathrm{i}k_x\sum_{\ell'}\psi^{\ell'}H_{\ell' m}\right] \\ &+ \mathcal{F}^{-1}_{k_\perp}[\mathrm{i}k_x J_{0s}\Phi_k\delta_{m0}]\mathcal{F}^{-1}_{k_\perp}\left[\mathrm{i}k_y\sum_{\ell'}\psi^{\ell'}H_{\ell' m}\right]\Bigg].\end{aligned} \tag{3.71}$$

Each term in the gyrokinetic equation has been projected onto the Hermite-Laguerre basis, with the exception of the nonlinear term. Collecting the time derivative (3.11), free streaming (3.27), curvature drift (3.32), $\nabla B$ drift (3.35), mirror (3.40), gradient drive (3.52), collisions (3.56), and nonlinear (3.93) terms, we have projected the gyrokinetic equation onto the Fourier-Laguerre-

Hermite basis:

$$\begin{aligned}
&\frac{\partial G_{\ell,m}}{\partial t} + (\sqrt{m+1}\varphi^{m+1} + \sqrt{m}\varphi^{m-1})\nabla_\parallel H_{\ell,m} \\
&\quad - [\ell\sqrt{m+1}H_{\ell-1,m-1} + \ell\sqrt{m+1}H_{\ell-1,m+1} \\
&\quad + (\ell+1)\sqrt{m+1}H_{\ell,m+1} + (\ell+1)\sqrt{m}H_{\ell,m-1}]\nabla_\parallel \ln B \\
&\quad + \mathrm{i}\omega_D^\kappa[\sqrt{m(m-1)}H_{\ell,m-2} + (2m+1)H_{\ell,m} + \sqrt{(m+1)(m+2)}H_{\ell,m+2}] \\
&\quad + \mathrm{i}\omega_d^{\nabla B}[(2\ell+1)H_{\ell,m} + (\ell+1)H_{\ell+1,m} + \ell H_{\ell-1,m}] \\
&\quad - \nabla_\parallel(\ln B)\sqrt{m}[(2\ell+1)H_{\ell,m-1} + \ell H_{\ell-1,m-1} + (\ell+1)H_{\ell+1,m-1}] \\
&\quad + \mathcal{N}_{\ell,m}^{GK} = \mathcal{D}_{\ell,m} + \mathcal{C}_{\ell,m}.
\end{aligned} \tag{3.72}$$

To close the system, we need a method to calculate $\Phi$ to compute the nonlinear and drive terms. We achieve this by projecting the quasineutrality condition (2.75) onto the Laguerre-Hermite basis. All plasmas in this thesis have a single ion species with $Z_i/e = 1$. After normalizing charges by $e$, the equilibrium density by itself, and temperatures by $T_i$, the quasineutrality condition becomes

$$\int \mathrm{d}^3\mathbf{v}\langle h\rangle_{\mathbf{r}} = \Phi + \frac{T_i}{T_e}[\Phi - \langle\langle\Phi\rangle\rangle]. \tag{3.73}$$

The projection of this expression onto the Laguerre-Hermite basis uses the Laguerre projection of the Bessel function for the gyroaverage (3.49), resulting in the following expression:

$$\sum_\ell \mathcal{J}_\ell H_{\ell,0} = \Phi + \frac{T_i}{T_e}[\Phi - \langle\langle\Phi\rangle\rangle]. \tag{3.74}$$

Equations (3.72) and (3.74) comprise the gyrokinetic system we solve using GX.

## 3.2 Vlasov-Poisson System

We solve the Vlasov-Poisson system (2.91) - (2.92) using an orthonormal spectral basis by projecting the velocity-space dependence onto a Hermite polynomial basis and the configuration-space dependence onto Fourier harmonics. We choose this approach because pseudo-spectral methods are often more efficient than pure finite-difference algorithms [Boyd, 2000], and as mentioned in the introduction, there is an extensive history of the application of the Fourier-Hermite basis to solve problems in kinetic plasma physics.

### 3.2.1 Hermite Moment Formalism

To discretize velocity space, we impose a Hermite basis representation, using notation similar to the notation from Mandell et al. [2018] that we used in (3.2). Including both the Maxwellian weight and the normalization factors, the Hermite basis expansion ($\varphi^m(v)$) and projection ($\varphi_m(v)$) functions are

$$\varphi^m(v) = \frac{\mathrm{He}_m(v)}{\sqrt{2\pi m!}} e^{-v^2/2}, \quad \varphi_m(v) = \frac{\mathrm{He}_m(v)}{\sqrt{m!}}, \tag{3.75}$$

where

$$\mathrm{He}_m(v) = (-1)^m e^{v^2/2} \frac{\mathrm{d}^m}{\mathrm{d}v^m} e^{-v^2/2} \tag{3.76}$$

are the probabilists' Hermite polynomials [Abramowitz and Stegun, 1972]. Due to the lower dimensionality of the Vlasov-Poisson system, we normalize the expansion function by a different factor in (3.75) than in (3.2). The different normalization yields this orthonormality relation:

$$\int_{-\infty}^{\infty} \mathrm{d}v \varphi^m(v) \varphi_{m'}(v) = \delta_{m,m'}, \tag{3.77}$$

Note that the normalized background Maxwellian in 1-D is identical to the $m = 0$ basis element:

$$\varphi^0 = \frac{1}{\sqrt{2\pi}} e^{-v^2/2} = F_M. \tag{3.78}$$

This property makes the Hermite moment representation a natural choice for near-Maxwellian velocity distributions. We project the fluctuating portion of the distribution function onto the Hermite basis using

$$g(z, v, t) = \sum_{m=0}^{\infty} \varphi^m G_m(z, t), \quad G_m(z, t) \equiv \int_{-\infty}^{\infty} \mathrm{d}v \varphi_m g(z, v, t). \tag{3.79}$$

### 3.2.2 Vlasov Equation in the Hermite basis

We seek an evolution equation for the amplitudes, $G_m$, of the Hermite expansion functions. To derive this moment hierarchy, we project (2.91) onto the Hermite basis term-by-term. The time derivative term is the most straightforward, as in electrostatic gyrokinetics:

$$\int_{-\infty}^{\infty} \mathrm{d}v \varphi_m \frac{\partial g}{\partial t} = \frac{\partial G_m}{\partial t}. \tag{3.80}$$

We then project the linear ballistic term, $v\,(\partial g/\partial z)$, onto the basis:

$$v\frac{\partial g}{\partial z} = \frac{\partial}{\partial z} \sum_{m=0}^{\infty} v\varphi^m G_m = \frac{\partial}{\partial z} \sum_{m=0}^{\infty} \left( \sqrt{m+1}\varphi^{m+1} + \sqrt{m}\varphi^{m-1} \right) G_m, \tag{3.81}$$

where we used the recurrence relation for the Hermite basis functions(3.23) to remove the explicit dependence on $v$:

$$\sqrt{m+1}\varphi^{m+1} = v\varphi^m - \sqrt{m}\varphi^{m-1}. \tag{3.82}$$

Applying the projection function to (3.81), and using (3.77) yields

$$\int_{-\infty}^{\infty} \mathrm{d}v \varphi_m v \frac{\partial g}{\partial z} = \frac{\partial}{\partial z}\left(\sqrt{m} G_{m-1} + \sqrt{m+1} G_{m+1}\right). \tag{3.83}$$

The linear portion of the electric field drive term contains the $m = 1$ expansion function, so applying the projection function to this term yields

$$\int_{-\infty}^{\infty} \mathrm{d}v \varphi_m v E F_M = E \int_{-\infty}^{\infty} \mathrm{d}v \varphi_m \varphi^1 = E \delta_{m,1}. \tag{3.84}$$

The final term to consider is the nonlinear term, $-E(\partial g/\partial v)$. Recall the derivative identity for the Hermite expansion functions (3.38):

$$\frac{\partial \varphi^m}{\partial v} = -\sqrt{m+1} \varphi^{m+1}. \tag{3.85}$$

The expansion of the nonlinear term in the Hermite basis is then

$$-E \frac{\partial g}{\partial v} = -E \sum_m \frac{\partial \varphi^m}{\partial v} G_m = E \sum_m \sqrt{m+1} \varphi^{m+1} G_m. \tag{3.86}$$

After projecting this term onto the basis,

$$\begin{aligned} \int_{-\infty}^{\infty} \mathrm{d}v \varphi_m (E \sum_{m'} \sqrt{m'+1} \varphi^{m'+1} G_{m'}) &= \sum_{m'} E \sqrt{m'+1} \delta_{m,m'+1} G_{m'} \\ &= E \sqrt{m} G_{m-1}, \end{aligned} \tag{3.87}$$

we can collect all of the terms to recover the general form of the equation for $G_m$:

$$\frac{\partial G_m}{\partial t} + \frac{\partial}{\partial z}\left(\sqrt{m}G_{m-1} + \sqrt{m+1}G_{m+1}\right) + E\sqrt{m}G_{m-1} + E\delta_{m,1} = 0. \tag{3.88}$$

### 3.2.3 Pseudo-spectral in Fourier Space

Next, we discretize space with a Fourier basis. The Fourier transform and inverse Fourier transform operators for a spatial grid of $N$ points are:

$$\mathcal{F}[g] = \frac{1}{N}\sum_z g\mathrm{e}^{-\mathrm{i}kz} \quad \mathcal{F}^{-1}[G_k] = \sum_k G_k \mathrm{e}^{\mathrm{i}kz}, \tag{3.89}$$

where the Fourier wavenumbers are given by $k = 2\pi j/N$, with $-N/2 \leq j < N/2$ and $j \in \mathbb{Z}$. The Fourier convolution theorem,

$$\mathcal{F}[fg] = \sum_{k'} F_k G_{k-k'}, \tag{3.90}$$

transforms the nonlinear term into a coupled sum over all wavenumbers. This coupling introduces additional computational overhead. We mitigate that problem by using a pseudo-spectral method [Boyd, 2000], evaluating the nonlinear term on the original configuration space and projecting the result back onto the space of Fourier wavenumbers. Applying the Fourier transform, (3.89), to (3.88) and then applying the Fourier convolution theorem yields

$$\frac{\partial G_{m,k}}{\partial t} + \mathrm{i}k\left(\sqrt{m}G_{m-1,k} + \sqrt{m+1}G_{m+1,k}\right) + \sqrt{m}\mathcal{F}[\mathcal{F}^{-1}[\mathbf{E}(k)]\mathcal{F}^{-1}[\mathbf{G}_{m-1}(k)]]_k = -E_k\delta_{m,1}, \tag{3.91}$$

where $\mathbf{E}(k)$ is the perturbed electric field in Fourier space, $\mathbf{G}_{m-1}(k)$ is a vector of all Fourier wavenumbers of the $(m-1)^{th}$ Hermite moment of the perturbed distribution function, and (3.89) defines the forward and inverse Fourier transforms.

Simulating this collisionless system with finite resolution without using any regularization is numerically infeasible, as the dynamics can generate structures in phase space at arbitrarily small scales [Parker and Dellar, 2015; Loureiro et al., 2016; Mandell et al., 2024]. To mitigate that effect, we introduce numerical regularization through hyperdiffusion $(-\nu_k k^4)$ and hypercollision $(-\nu_m m^4)$ operators:

$$H_{mk} = \begin{cases} 0 & \text{if } m \leq 2 \\ -\nu_k k^4 - \nu_m m^4 & \text{if } m > 2, \end{cases} \tag{3.92}$$

where $\nu_k$ and $\nu_m$ are parameters which control the strengths of the hyperdiffusion and hypercollisions, respectively. Note that we do not apply this regularization to the $m = 0$, 1, and 2 equations, as they encode conservation of particle number, momentum, and energy, respectively.

We solve the system for a finite number of moments $M$. We introduce the following compact notation for the $k^{th}$ Fourier wavenumber of the nonlinear operator, $\mathcal{N}_k$:

$$\mathcal{N}[\mathbf{E}(k), \mathbf{G}_{m-1}(k)]_k = \sqrt{m}\mathcal{F}[\mathcal{F}^{-1}[\mathbf{E}(k)]\mathcal{F}^{-1}[\mathbf{G}_{m-1}(k)]]_k. \tag{3.93}$$

Accounting for the special cases of the $m = 0$, $m = 1$, and $m = M-1$ equations, we have derived

the projection of the Vlasov equation onto the Fourier-Hermite basis:

$$\frac{\partial G_{0,k}}{\partial t} + \mathrm{i}kG_{1,k} = 0, \tag{3.94}$$

$$\frac{\partial G_{1,k}}{\partial t} + \mathrm{i}k\left(G_{0,k} + \sqrt{2}G_{2,k}\right) + \mathcal{N}[\mathbf{E}(k), \mathbf{G}_0(k)]_k + E_k = 0, \tag{3.95}$$

$$\frac{\partial G_{m,k}}{\partial t} + \mathrm{i}k\left(\sqrt{m}G_{m-1,k} + \sqrt{m+1}G_{m+1,k}\right) + \mathcal{N}[\mathbf{E}(k), \mathbf{G}_{m-1}(k)]_k = H_{m,k}G_{m,k}, \quad 2 \leq m < M-1, \tag{3.96}$$

$$\frac{\partial G_{M-1,k}}{\partial t} + \mathrm{i}k\sqrt{M-1}G_{M-2,k} + \mathcal{N}[\mathbf{E}(k), \mathbf{G}_{M-2}(k)]_k = H_{M-1,k}G_{M-1,k}. \tag{3.97}$$

### 3.2.4 Poisson's Equation in the Fourier-Hermite basis

To complete the system, we project Poisson's equation onto the Fourier-Hermite basis and solve for the electric field, $E_k$. First, we solve for the charge density in the Hermite basis:

$$\rho = -\int_v \mathrm{d}v g = -\int_v \mathrm{d}v \varphi_0 \sum_m \varphi^m G_m = -G_0. \tag{3.98}$$

Next, we insert this expression for the normalized charge density into Poisson's equation (2.92) and apply the Fourier transform (3.89) to calculate the electric field as a function of the $m = 0$ density moment in the Fourier-Hermite basis:

$$E_k = \frac{\mathrm{i}G_{0,k}}{k}. \tag{3.99}$$

We require that the $k = 0$ mode is zero and stationary, corresponding to zero mean electric field.

## 3.3 Summary

In this chapter, we discussed the merits of different approaches to the numerical solutions of PDEs, and we motivated our decision to use Eulerian, pseudo-spectral methods. We then derived the projection of the electrostatic gyrokinetic system onto a Fourier-Laguerre-Hermite basis and the Vlasov-Poisson system onto a Fourier-Hermite basis. This was our attempt to answer the broad question, "What method should we use to represent the domain and differential operators?" More questions remain, including the question, "How much resolution should we use to represent the domain?" Decreasing resolution often decreases the fidelity of simulations, while also reducing their computational expense. In the next chapter, we leverage data-driven, machine-learning (ML) methods to reduce the velocity resolution requirements for the Vlasov-Poisson system, pointing toward a promising path to accelerate kinetic simulations of plasmas.

# Chapter 4: Machine-learning Closure of Vlasov-Poisson Dynamics in Fourier-Hermite Space

Much of this chapter was orginially published in the *Journal of Plasma Physics* as a paper with the same name as the title of the chapter [Barbour et al., 2025]. This work was conducted in collaboration with Prof. William Dorland, Dr. Ian G. Abel, and Dr. Matt Landreman and has been reused with the permission of the surviving authors.

## 4.1 Motivation

One of the most important ongoing efforts in the pursuit of magnetic fusion energy is the study of microturbulence [Yoshida et al., 2025]. While magnetic-confinement fusion plasmas are designed to be stable to macroscopic magnetohydrodynamic (MHD) perturbations, the steep gradients in temperature and density that are present in all current designs also drive small-scale instabilities. These instabilities drive sustained turbulence, which then dominates the transport of particles, energy, and momentum in the plasma.

As the fusion industry turns its focus toward designing first-of-a-kind pilot plants, optimizing these designs to mitigate turbulent heat losses is an urgent objective. Significant progress in turbulence optimization for tokamaks [Highcock et al., 2018] and stellarators [Roberg-Clark et al., 2023; Kim et al., 2024] has been achieved. However, the space of candidate magnetic equilibria remains vast, and high-resolution turbulence simulations are computationally expensive. A unifying feature of fluid [Kolmogorov, 1991; Kraichnan, 1959], MHD [Goldreich and Sridhar, 1995], and gyrokinetic [Bañón Navarro et al., 2011] turbulence theory is that free energy tends to cascade

from large scales to small scales through local interactions in phase space, and these cascades are well-described by power laws[1]. Therefore, models that can accurately calculate large-scale quantities, including spectra and heat fluxes, without needing to resolve small scales, are both desirable and plausible.

In the field of fluid turbulence modeling, two successful approaches to constructing reduced, lower-resolution models that capture large-scale behavior are Reynolds-Averaged Navier-Stokes (RANS, Reynolds [1895]) and large eddy simulation (LES, Deardorff [1970]). LES also has been used to build reduced models of gyrokinetic turbulence [Bañón Navarro et al., 2014]. These formalisms accelerate turbulence simulations by averaging over time and filtering out the small-scale dynamics. Closure models provide another pathway to filter out the small scales while retaining accurate large-scale statistics. In kinetic plasma physics, Hammett and Perkins [1990] developed a landmark analytic linear closure for the low-frequency response to small Langmuir oscillations.

An alternative method for building reduced models is to use data-driven, statistical techniques like machine learning (ML) and Bayesian optimization. In recent years, the proliferation of ML models has led to advancements in pattern identification and model construction for magnetic fusion. Some example applications include improved prediction and avoidance of tokamak disruptions using the `DECAF` framework [Piccione et al., 2020] and on the DIII-D tokamak [Rea et al., 2019; Fu et al., 2020], improved active-feedback plasma control systems with reinforcement learning [Degrave et al., 2022; Seo et al., 2024; Kerboua-Benlarbi et al., 2024], optimized execution of gyrokinetic simulations with the `PORTALS` framework [Rodriguez-Fernandez et al., 2022],

[1]Power laws typically arise when a scale-local transfer of energy occurs, and a separation between energy injection and dissipation scales exists. The relevance of scale locality to energy cascades in gyrokinetics has been studied in detail by Barnes et al. [2011] and Teaca et al. [2017]. In some systems of interest, inverse cascades occur, driving invariants from small scales to large scales. The theory of inverse cascades has been developed for several two-dimensional systems [Hasegawa and Mima, 1978; Plunk et al., 2010], yet an open question remains about their applicability to the full five-dimensional gyrokinetic system.

and accelerated, low-resolution turbulence simulations [Greif et al., 2023; Castagna et al., 2024; Clavier et al., 2025]. Data-driven, configuration space [Ma et al., 2020; Qin et al., 2023] and heat flux [Ingelsten et al., 2024; Huang et al., 2025] closure models of Landau damping have also been developed.

In this chapter, we choose to use the one-dimensional, collisionless Vlasov-Poisson system as a test problem for implementing an ML closure in velocity space. This system is the one-dimensional, electrostatic limit of the full five-dimensional gyrokinetic system [Parker and Dellar, 2015]. It models dynamics parallel to the magnetic field, and it contains a quadratic nonlinearity present in the full gyrokinetic system. These features make the Vlasov-Poisson system a suitable foundation to assess the future potential for a closure model to be developed for gyrokinetic turbulence.

In parallel velocity space, we use properties of the Hermite basis to cast the objective of reducing resolution requirements as a closure problem. The Hermite basis has been used extensively to solve problems in kinetic plasma dynamics [Armstrong, 1967; Grant and Feix, 1967; Joyce et al., 1971; Gibelli and Shizgal, 2006; Zocco and Schekochihin, 2011; Loureiro et al., 2013; Parker and Dellar, 2015; Kanekar et al., 2015; Jorge et al., 2017; Adkins and Schekochihin, 2018; Mandell et al., 2018; Frei et al., 2023]. In this work, we opt to focus on velocity space because the Hermite basis naturally expresses the interaction between scales in velocity space as local interactions in a coupled hierarchy of moments. This property of the moment formalism directly presents a single-term target for a closure. Specifically, the interaction between resolved scales and unresolved scales in velocity space appears as a single unknown term in an equation in this representation. In contrast, the spectral representation of configuration space introduces some nonlocal interactions across scales, complicating the development of a closure model. Additionally, the Hermite

basis explicitly expresses conservation laws for particle number, momentum, and energy as evolution equations for the lowest-order moments. These properties allow us to develop a closure that preserves the low-order conservation laws. More details on the closure problem are presented in Section 4.1.1 .

We use reservoir computing, an ML architecture, to implement a spectral, proof-of-concept, velocity-space closure for the moment hierarchy that maintains accurate spectra and reduces resolution requirements. The reservoir computing model maintains an internal representation of the system state, which it evolves through a directed graph with fixed weights. It then predicts the system state at time $t+1$ from the state at time $t$ using a set of output weights trained by linear regression. Since its concurrent development by Jaeger [2001] and Maass et al. [2002], reservoir computing has successfully been applied to the task of forecasting the dynamics of low-dimensional chaotic systems [Lu et al., 2017; Pathak et al., 2018]. This ML paradigm is well-suited to the task of modeling time-series data of physical processes. In climate physics, Arcomano et al. [2020] built a surrogate global atmospheric forecast model with reservoir computing, and in plasma physics, Jalalvand et al. [2022] used reservoir computing to analyze and classify Alfvén eigenmodes using DIII-D diagnostics data. In comparison to several other recurrent neural network architectures, reservoir computing has shown comparable ability to capture the statistics of chaotic systems, while requiring significantly less training time [Vlachas et al., 2020]. These results motivate our decision to use reservoir computing for the closure model.

Some ML methods, including reservoir computing, are more challenging to interpret than traditional physics models. Artificial neural network architectures are often "black boxes" that generate layered, analytically-intractable networks of nonlinear activation functions. Gaining insightful intuition from these networks can be daunting. We mitigate this problem by constraining

the ML portion of our simulations to represent only the unresolved small scales, preserving analytic evolution equations for large scales. Our methods are analogous to an LES approach for phase space.

This chapter has the following structure. In Section 4.1.1, we present the closure problem. We introduce our machine-learning closure model in Section 4.2 as a solution to that problem. In Section 4.3 , we present our ML model's predictions, and we conclude with a summary and discussion in Section 4.4 . Our test problem is the one-dimensional Vlasov-Poisson system, introduced in Section 2.3. We use the Fourier-Hermite pseudo-spectral method derived in Section **??** to solve the coupled system of equations.

### 4.1.1 The Closure Problem

A continuous representation formally requires an infinite number of Hermite moments, but we must discretize the system with a finite number of moments, $M$. Observe that in the Fourier-Hermite basis, the parallel streaming term in the convective derivative, $v(\partial g/\partial z)$, becomes $\mathrm{i}k\left(\sqrt{m}G_{m-1,k} + \sqrt{m+1}G_{m+1,k}\right)$ for general $m$. This term couples nearest-neighbors in Hermite space and forms a hierarchy of equations that must be closed. This closure problem is analogous to the classic closure problem encountered when taking fluid moments of the kinetic equation (2.81). The resulting fluid equations for density, momentum, energy, and higher order moments exhibit coupling between moment $m$ and moments $m+1$ and $m-1$. This same feature occurs in the Hermite representation of velocity space, and from (3.79) and (3.75), the Hermite moment basis is obtained through a similar process of multiplying the distribution function by increasing powers of $v$ via increasing orders of the Hermite polynomials and then integrating over velocity space.

The simplest model of closure is by truncation, which we use in the direct numerical solution by omitting the $G_{m+1,k}$ term in the evolution equation for the final moment, $G_{M-1,k}$. Our objective is to introduce a dynamical closure model to close the hierarchy at a moment number $m = m_c$, $m_c < M - 1$, while still achieving accurate predictions of the Fourier and Hermite spectra. We use a reservoir recurrent neural network as the closure model, which we describe in more detail in Section 4.2 .

## 4.2 Machine-Learning Closure: Reservoir Computing

Inspecting the evolution equation for $G_{m,k}$ (3.88) reveals that the coupling between moment $m$ and moment $m + 1$ only explicitly appears as a linear term in the equation, acting on each wavenumber $k$ independently. As a result, we introduce a unique and independent Hermite closure model for each $k$ in the system, for a total of $N_k$ reservoirs. We use methods similar to Pathak et al. [2018] to design the reservoirs.

Each reservoir is a network of $D_r$ nodes defined by a latent state vector $\boldsymbol{r}_k(t) \in \mathbb{R}^{D_r}$, where each element of $\boldsymbol{r}_k(t)$ represents the state of one node at time $t$. The structure of the reservoir computing model is presented in Figure 4.1 . The connections between nodes are defined by a $D_r \times D_r$ sparse adjacency matrix, $\boldsymbol{A}$, where exactly $\kappa$ nonzero elements per row are initialized to random values drawn from the uniform distribution on [0,1] and scaled by the appropriate factors to set the largest eigenvalue of $\boldsymbol{A}$ equal to a specified value, $\rho_{sp}$. The magnitude of the largest eigenvalue of a reservoir's adjacency matrix has been demonstrated to significantly impact the accuracy of the reservoir's predictions [Jiang and Lai, 2019]. The runtime is separated into two phases, training and prediction. In the training phase, the inputs are derived from a high-resolution

direct numerical solution (DNS). In the subsequent prediction phase, the inputs are derived from a low-resolution simulation with the output of our reservoir used to compute the closure. The input to a given reservoir at each time step is a vector $\boldsymbol{u}_k(t)$ of dimension $2w$, where the elements of $\boldsymbol{u}$ are the real and imaginary parts of the final $w$ Hermite moments before $m_c$:

$$\boldsymbol{u}(t) = \left[G_{m_c-w,k}(t - w\Delta t),\, G_{m_c-(w-1),k}(t - (w-1)\Delta t),\, \ldots G_{m_c-1,k}(t - \Delta t)\right]. \tag{4.1}$$

In both phases, the inputs form a diagonal in $m$ and $t$ of Hermite moments, and this structure is inspired by the phenomenon of linear Landau damping in the linearized Vlasov-Poisson system, where energy cascades from low $m$ to high $m$ in time. The input layer is a $D_r \times 2w$ matrix $\boldsymbol{W}_{in}$, with its elements drawn from the uniform distribution on $[-\sigma, \sigma]$, where $\sigma$ is a parameter. Critically, $\boldsymbol{A}$ and $\boldsymbol{W}_{in}$ remain fixed after initialization. Only the output layer, $\boldsymbol{W}_{out}$, is trained. The reservoir state vector evolves with the equation

$$\boldsymbol{r}_k(t + \Delta t) = \tanh[\boldsymbol{A}\boldsymbol{r}_k(t) + \boldsymbol{W}_{in}\boldsymbol{u}(t)], \tag{4.2}$$

where the hyperbolic tangent activation function operates element-wise on the vector argument.

The elements of the reservoir-to-output coupling, $\boldsymbol{W}_{\text{out}}$, are parameters that must be tuned to train the reservoir. We introduce a hyperparameter $b$ to exclude data from initial transients from the reservoir training process. This hyperparameter serves as a buffer of time steps between the initialization time and the data assigned for training $\boldsymbol{W}_{\text{out}}$. During the training phase, defined by the times $t$ where $b\Delta t \leq t \leq T$, the reservoir output is trained to approximate the closure moment, $G_{m_c,k}$.

The most straightforward choice for the closure evaluation would be the linear readout operation, $\boldsymbol{W}_{out}\boldsymbol{r}_k$. Empirically, that operation does not permit reservoirs to predict the dynamics of several other nonlinear systems, including the Lorenz system [Lu et al., 2017; Chattopadhyay et al., 2020; Pyle et al., 2021], Kuramoto-Sivashinsky equation [Pathak et al., 2018], and the global atmospheric system [Arcomano et al., 2020]. As Lu et al. [2017] and Pathak et al. [2018] describe, the hyperbolic tangent operation in the evolution equation for the reservoir state (4.2) introduces an odd-parity symmetry to the reservoir state vector $\boldsymbol{r}$. As in the Kuramoto-Sivashinsky equation and the Lorenz systems studied by these authors, the quadratic nonlinear term in the Vlasov equation breaks odd-parity symmetry. In other words, the general evolution equation for the Hermite moments (3.91) is *not* invariant under the transformation $G_{m,k} \to -G_{m,k}$, while the evolution equation for the reservoir state (4.2) *is* invariant under the transformations $\boldsymbol{u} \to -\boldsymbol{u}$ and $\boldsymbol{r} \to -\boldsymbol{r}$. Following the methods of Pathak et al. [2018] and explored in more detail by Chattopadhyay et al. [2020] and Pyle et al. [2021], we resolve this conflict by applying a symmetry-breaking transformation ($\boldsymbol{r}_k \to \boldsymbol{r}_k^*$) to the reservoir latent state before readout:

$$r_{i,k}^*(t) = \begin{cases} r_{i,k}(t), & \text{even } i \\ \\ r_{i,k}^2(t), & \text{odd } i. \end{cases} \tag{4.3}$$

Because the reservoir output, $\boldsymbol{W}_{out}\boldsymbol{r}^*(t)$, is a linear operation on $\boldsymbol{r}^*$, we can train the weights of $\boldsymbol{W}_{out}$ using linear regression and avoid the computational expense of backpropagation. To mitigate potential overfitting, we introduce a regularization parameter, $\beta$, in the method of Tikhonov-regularized linear regression. The training problem is then to find the values of $\boldsymbol{W}_{out}$ which

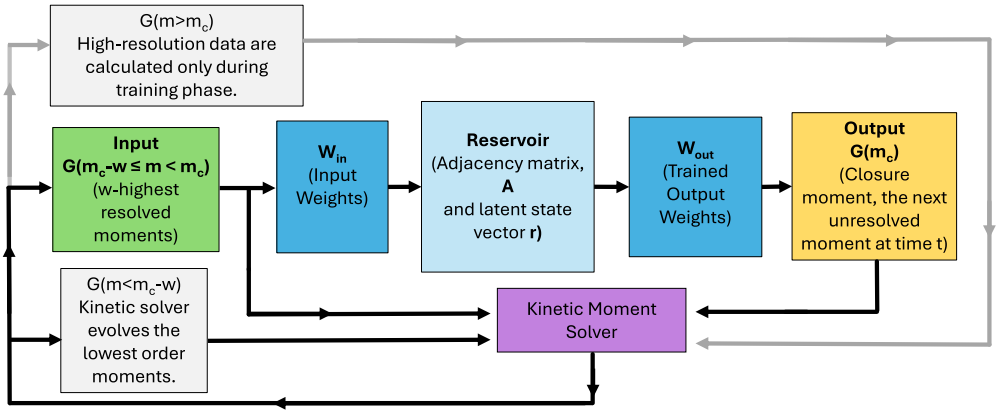



Figure 4.1: Diagram of the reservoir computing closure model, integrated with the kinetic moment solver.

minimize

$$\sum_{j=b}^{T/\Delta t} ||G_{m_c,k}(j\Delta t) - \boldsymbol{W}_{\text{out}}\boldsymbol{r}_k^*(j\Delta t)||^2 + \beta||\boldsymbol{W}_{out}||^2. \tag{4.4}$$

Explicitly, the closed-form solution that solves this linear regression problem is

$$\boldsymbol{W}_{out} = \boldsymbol{G}_{m_c,k}\boldsymbol{R}_k^{*T}(\boldsymbol{R}_k^*\boldsymbol{R}_k^{*T} + \beta\boldsymbol{I})^{-1}, \tag{4.5}$$

where $\boldsymbol{G}_{m_c,k}$ is a $2 \times (T/\Delta t - b)$ matrix consisting of the time series of the real and imaginary parts of the closure moment, $G_{m_c,k}$, $\boldsymbol{R}_k^*$ is a $D_r \times (T/\Delta t - b)$ matrix consisting of the time series of the modified state vector $\boldsymbol{r}_k^*(t)$, $\boldsymbol{R}_k^{*T}$ is the transpose of $\boldsymbol{R}_k^*$, and $\boldsymbol{I}$ is the identity matrix.

### 4.2.1 Closure Integration

Once the weights of $\boldsymbol{W}_{out}$ are trained, the algorithm switches to the prediction phase, and the system does not evolve moments beyond $m = m_c$. Instead, the output of the reservoir is fed back into the system of equations as a closure for the moment hierarchy. To make the discussion

more concrete, we introduce a time integration operator $\mathcal{D}$, such that $\mathcal{D}[\boldsymbol{G}(t)] = \boldsymbol{G}(t + \Delta t)$, where $\boldsymbol{G}(t)$ is the $M \times N_k$ matrix of Fourier-Hermite amplitudes representing the state of the system at time $t$. $\mathcal{D}$ is a generic operator, and its implementation details are flexible. For the results presented in this chapter, we implement $\mathcal{D}$ using the Python libraries `NumPy` [Harris et al., 2020] and `SciPy` [Virtanen et al., 2020] with the classic fourth-order explicit Runge-Kutta method (RK4) and the third-order Strong Stability-Preserving Runge-Kutta method (SSPRK3) as detailed by Durran [2010].

When we introduce the closure models, the time integration becomes a two-step process. For each wavenumber $k$, we use the time integration operator $\mathcal{D}$, to calculate the resolved moments $\boldsymbol{G}_{m<m_c,k}$:

$$\boldsymbol{G}_{m<m_c,k}(t + \Delta t) = \mathcal{D}[\boldsymbol{G}_{m\leq m_c}(t)]_k. \tag{4.6}$$

We then use a reservoir to predict the closure moment $G_{m_c,k}$:

$$G_{m_c,k}(t + \Delta t) = \boldsymbol{W}_{out}\boldsymbol{r}^*_k(t + \Delta t). \tag{4.7}$$

### 4.2.2 Computational Complexity

Though performance profiling metrics for numerical solutions of partial differential equations are highly dependent on computational hardware architectures and algorithm implementation choices, we can calculate the computational complexity of the DNS and the ML closure for the cases in this chapter. Both methods have an identical cost to evolve the lowest-order moments. Thus, we only report the cost of evolving moments $m_c \leq m < M$ for the DNS and the closure moment $m_c$ for the ML method. While the time-integration operator $\mathcal{D}$ is generic, we will remove

a layer of abstraction by calculating the result for the SSPRK3 algorithm. For simplicity, we will neglect the additional cost of operations on complex numbers relative to those same operations on real numbers, with the understanding that this will underestimate the cost of the DNS in comparison to the ML closure. The cost of transcendental functions including $\tanh$ depends on instruction sets. This estimate treats them as equivalent in cost to multiplication, with the understanding that this underestimates the cost of updating the reservoir state. We also neglect the cost of memory access and copy operations for both methods.

#### 4.2.2.1 Complexity of DNS

First, let us introduce a new parameter $M' = M - m_c$ that represents the additional number of moments that the DNS must evolve (3.96) in comparison to the ML closure model. Our choice of time-integration operator, SSPRK3, is a three-stage algorithm. Each stage includes an evaluation of the electric field, nonlinear term, linear streaming term, and hypercollisions. The electric field is calculated by (3.99). The vector $(\mathrm{i}/k)$ can be stored in memory during initialization, so the electric field calculation requires $N_k$ multiplication operations. The nonlinear term is evaluated with a pseudo-spectral method using (3.93). Using the $\mathcal{O}(N \log N)$ scaling of the fast Fourier Transform (FFT) algorithm [Cooley and Tukey, 1965], the total computational complexity of the nonlinear term is $3M'\mathcal{O}(N_k \log N_k)$. If the vector $(\mathrm{i}k)$ is also stored during initialization, then the linear streaming term requires $6M'N_k$ operations. The dissipation matrix in 3.92 can be stored, and the cost of the regularization term is $2M'N_k$. Finally, we include the operations needed to combine the results of each stage into the state of the system at the next time step. For SSPRK3, this requires

$12M'N_k$ operations. The total cost of DNS for each time step is then

$$\text{Cost}_{\text{DNS}} = 3M'\mathcal{O}(N_k \log N_k) + 20M'N_k + N_k. \tag{4.8}$$

#### 4.2.2.2 Complexity of ML Closure

The ML closure evaluation has two stages, updating the latent state of each reservoir and evaluating the closure moment. There are a total of $N_k$ reservoirs. The state of each reservoir, $\boldsymbol{r}_k$, updates at every time step using (4.2). This requires the sum of two matrix-vector products. The product $\boldsymbol{W}_{in}\boldsymbol{u}$ has a cost of $4D_r w$ operations. The other product $\boldsymbol{A}\boldsymbol{r}_k$ involves a sparse matrix. Though the cost of this product is implementation-specific, we will count operations on nonzero elements, resulting in a cost of $2D_r\kappa$ operations, where $\kappa < D_r$. Then, the $\tanh$ operation is applied to the sum of these products. The total cost of the reservoir update is then $2D_r[\kappa + 2w + 1]$. The symmetry-breaking transformation (4.3) is applied to the state vector, which requires $D_r/2$ multiplication operations. Finally, each reservoir predicts a Hermite closure using (4.7), which uses $4D_r$ operations. The ML closure is linear in $k$. In total, the cost calculating the ML closure for each time step is then

$$\text{Cost}_{\text{ML}} = N_k D_r[8w + 4\kappa + 13/2]. \tag{4.9}$$

After eliminating the common factor of $N_k$ from the cost of the DNS and the ML closure, we find that as $N_k$ increases, the dominant scaling for the DNS is $\mathcal{O}(M' \log N_k)$. In contrast, the cost of the ML closure scales linearly with the size of each reservoir $D_r$. For sufficiently large $(M' \log N_k)/(D_r w)$, the reservoir closure will be faster than the DNS.

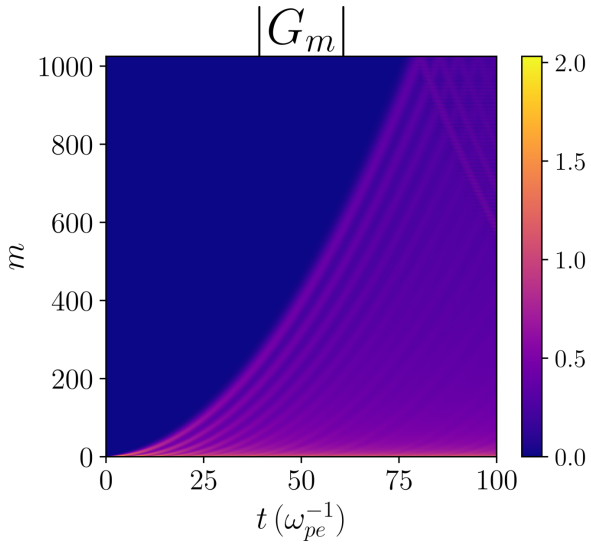



Figure 4.2: High-velocity-resolution ($M = 1025$) baseline time series of Hermite amplitudes for the driven, linearized Vlasov-Poisson system. Energy is injected as a density perturbation in the $m = 0$ Hermite moment and advects to higher moments through linear Landau damping. No hypercollisional regularization is applied to this example, and hundreds of Hermite moments are required to prevent numerical reflection at the high-$m$ boundary from impacting the low-$m$ spectrum.

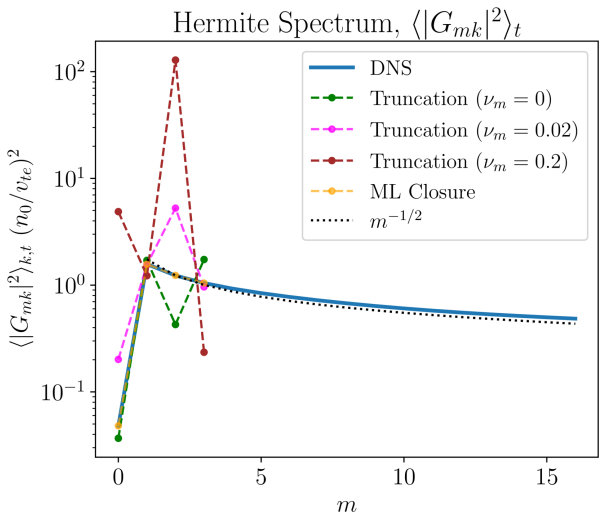



Figure 4.3: Time-averaged Hermite spectra for the ML closure model compared to the high-velocity-resolution ($M = 1025$) baseline, closure-by-truncation with three different coefficients ($\nu_m$) for hypercollisional regularization, and the theoretical $m^{-1/2}$ scaling. The ML closure model shows strong agreement with the baseline simulation, while no value of $\nu_m$ produces an accurate spectrum.

## 4.3 Closure Results

### 4.3.1 Linearized Vlasov-Poisson

For our first test of the ML closure, we confirm that the model accurately solves the linear limit of the Vlasov-Poisson system. First, we linearize the Vlasov equation (2.91), dropping the

nonlinear term:

$$\frac{\partial g}{\partial t} + v\frac{\partial g}{\partial z} + vEF_M = 0. \tag{4.10}$$

In Fourier space, this allows us to restrict our focus to a single wavenumber, as the coupling between wavenumbers occurs only in the nonlinear term. We simulate an approximately steady-state solution by driving the system, as explored by Kanekar et al. [2015]. We inject energy into the $m = 0$ moment, and it cascades down to finer scales through Landau damping. We incorporate these features into the projection of this equation onto the Fourier-Hermite basis using the results of Section **??** :

$$\frac{\partial G_{m,k}}{\partial t} + \mathrm{i}k\left(\sqrt{m}G_{m-1,k} + \sqrt{m+1}G_{m+1,k}\right) = \chi(t)\delta_{m,2} - E_k\delta_{m,1}, \tag{4.11}$$

where $\chi(t)$ is a forcing coefficient independently drawn at each time step from the uniform distribution on $[0, 1)$. The electric field $E_k$ is given by (3.99). The purpose of this mechanism for energy injection is to achieve a statistical steady-state solution. Our primary systems of interest exhibit nonlinear chaos and typically settle into time-dependent steady states. While constant forcing would also achieve a steady-state solution, a time-dependent forcing mechanism more closely resembles typical problems encountered in plasmas. Other injection mechanisms, including driving the $m = 2$ moment or applying a stochastic external electric field yield similar performance for the closure model. Figure 4.2 depicts a high-resolution ($M = 1025$) baseline simulation of (4.11) with $k = 0.4$.

When no hypercollisional regularization is applied, the cascade of free energy numerically reflects at the high-$m$ boundary. In the simulation in Figure 4.2, this reflection occurs around

$t = 80$. This phenomenon always occurs at a finite time for a given Hermite resolution ($M$), meaning that extending the duration of a simulation where the low-order moments are valid requires increasing the Hermite resolution. Additionally, the rate at which energy advects from low-$m$ to high-$m$ increases with Fourier wavenumber $k$ [Parker and Dellar, 2015]. Therefore, though this linearized simulation with $k = 0.4$ is feasible with high velocity resolution, more Hermite resolution would be required for a nonlinear simulation with moderate spatial resolution. Some form of artificial dissipation is necessary for baseline simulations of higher-dimensional systems to be computationally feasible. We therefore apply hypercollisional regularization to the nonlinear form of the Vlasov-Poisson system in Sections 4.3.2 and 4.3.3 .

We then train a reservoir to predict the Hermite closure moment, $G_{m_c,k}$, where we evaluate the closure at $m_c = 3$. The choice of $m_c = 3$ is motivated by the tradition of heat flux closures in fluid theory and our desire for a closure that allows the conservation laws in the $m \in [0, 1, 2]$ equations to remain intact. One motivation for a closure is to reduce the resolution required to achieve accurate simulations, so we choose the lowest valid value for $m_c$. We use the time series of the density, momentum, and temperature moments from the time window $15 \leq t < 25$ as training data. We find that a requirement is that enough time has passed to allow energy to cascade into $m = m_c$. We report results training on data from the time window $15 < t \leq 25$, but the reservoir supports accurate spectra for other training windows as well. Figure 4.3 depicts the excellent agreement between the low-order spectra calculated by the high-resolution baseline and the ML closure model. The mean-squared error in the ML closure spectrum is $2.25 \times 10^{-5}$. The ML closure shows favorable performance in comparison to a naïve closure by truncation, where no hypercollisional regularization is applied. It also outperforms a closure by truncation where the hypercollisional term, $-\nu_m m^4 G_{m,k}$, is introduced to the right-hand side of (4.11) to mitigate

numerical reflection at the high-$m$ boundary. As in (3.92), this term is only applied for $m > 2$ to preserve the low-order conservation laws. For an initial comparison, $\nu_m$ is set to $2 \times 10^{-2}$ to maintain finite dissipation at $m_c = 3$, where $\nu_m m_c^4$ is order one. For reference, we also compare to the theoretical scaling $G_m \sim m^{-1/2}$ [Zocco and Schekochihin, 2011; Kanekar et al., 2015].

The reservoir hyperparameters used to construct the model are spectral radius $\rho_{sp} = 0.6$, adjacency matrix degree of three, Tikhonov regularization parameter $\beta = 10^{-9}$, input scaling parameter $\sigma = 0.5$, and six nodes in the reservoir. We found these hyperparameters by comparing mean-squared errors in spectra. An unguided scan led to a reduction in error from $1.75 \times 10^{-7}$ to $2.26 \times 10^{-8}$. No systematic optimization was required, though some choices of parameters led to numerically unstable time integration of (4.11). We used the same procedure to select the hyperparameters in both the weakly and strongly nonlinear cases presented in Sections 4.3.2 and 4.3.3, respectively. The spectral radius $\rho_{sp}$ of the adjacency matrix $\boldsymbol{A}$ must be set to $\rho_{sp} < 1$ for stability. The block-diagonal structure for $\boldsymbol{W}_{in}$ that Vlachas et al. [2020] used in their reservoir computing implementation yielded unstable predictions. We speculate that this may be related to differences in requirements for complex-valued data in comparison to real-valued data, but more analysis would be required to confirm. For the strongly nonlinear case in Section 4.3.3 , using a reservoir smaller than sixty nodes or including the initial transient in the training set also led to unstable solutions. The computational complexity for the evaluating the ML closure is the same as in 4.9 , with $N_k = 1$. The DNS cost for this case is $18M' + 1$, after removing the contributions of the nonlinear term and hypercollisions from (4.8).

### 4.3.2 Weakly Nonlinear Vlasov-Poisson

Next, we reintroduce the nonlinear term, $-E(\partial g/\partial v)$, which permits coupling between wavenumbers. This coupling in Fourier space allows energy to flow across spatial scales, in addition to the already-present energy cascade in velocity space. This effect is subdominant at low amplitudes, but it becomes more significant as amplitudes increase. We first examine the low-amplitude limit in this section before proceeding to a high-amplitude case in Section 4.3.3 . In the low-amplitude limit, Landau damping continues to be the dominant effect. Here, we solve an initial value problem, where the initial condition is a cosine density perturbation with a Maxwellian velocity component:

$$g(z, v, t = 0) = \epsilon \cos(k_0 z) F_M(v), \tag{4.12}$$

where $\epsilon$ is the initial amplitude of the perturbation normalized by the background density and Maxwellian weight and $k_0$ is the wavenumber of the initial perturbation. In the spectral domain, this initial condition has the convenient form

$$G_{m,k,t=0} = \frac{\epsilon}{2} \delta_{m,0} \left( \delta_{k,k_0} + \delta_{k,-k_0} \right), \tag{4.13}$$

collapsing the initial system state into a single nonzero value after applying the reality condition. To explore the linear limit of the system, we choose $\epsilon = 0.001$, a $0.1\%$ density perturbation. We examine the $k_0 = 0.4$ case that Brunetti et al. [2000] explored in their analysis, accounting for a factor of ten difference in normalizations. With this low initial amplitude, we choose $N_k = 17$ Fourier wavenumbers, including the $k = 0$ mode, anticipating that the dominant energy cascade will be linear Landau damping. For the high-resolution simulations, we continue to use $M = 17$

Hermite moments, including the $m = 0$ moment. The hypercollisional regularization coefficient is $\nu_m = 5 \times 10^{-5}$, and we set $\nu_k = 0$. This value of $\nu_m$ is chosen to maintain finite dissipation at the finest velocity scale in the simulation, such that the magnitude of the coefficient of the hypercollisional term at that scale, $\nu_m(M-1)^4$, is order one. As the Fourier spectrum in Figure 4.6 shows, the flux of energy to high $k$ in this regime is small, so $\nu_k$ is not necessary. We set the size of the spatial simulation domain to $L = 5\pi$ so that $1/k_0$ is the largest wavelength resolved in the system. Figure 4.4 presents a comparison between the baseline, high-Hermite-resolution simulation, the ML closure model, and the theoretical damping rate. The time trace for the ML closure method begins at the end of the training phase, $t = 25$. To calculate the damping rate, we use the complex root-finding algorithm of Carpentier and Dos Santos [1982], implemented in the generalized dispersion relation solver developed by Ivanov and Adkins [2023]. These results demonstrate that with a low initial amplitude, our baseline simulation converges to the expected linear Landau damping rate. Additionally, Figure 4.4 demonstrates that the ML closure model preserves the frequency of the wave. Time traces of the $m = 1$ and $m = 2$ moments show similar performance and are presented in Appendix A.2 .

In Figure 4.5 , we compare the ML closure model, with $m_c = 3$, to both the high-resolution ($M = 17$) baseline and two low-resolution simulations without the ML closure, i.e. closure by truncation. We also compare the Fourier spectra of these simulations in Figure 4.6 . For this problem, we place an independent reservoir at each wavenumber $k$ and train it to learn a Hermite moment closure for that wavenumber. The ML closure shows strong agreement with the high-resolution baseline, with mean-squared errors of $1.26 \times 10^{-42}$ for the Hermite spectrum and $8.01 \times 10^{-38}$ for the Fourier spectrum. It successfully captures the Hermite and Fourier spectra of the system and confirms that its predictive capability continues to hold when a small, but nonzero,

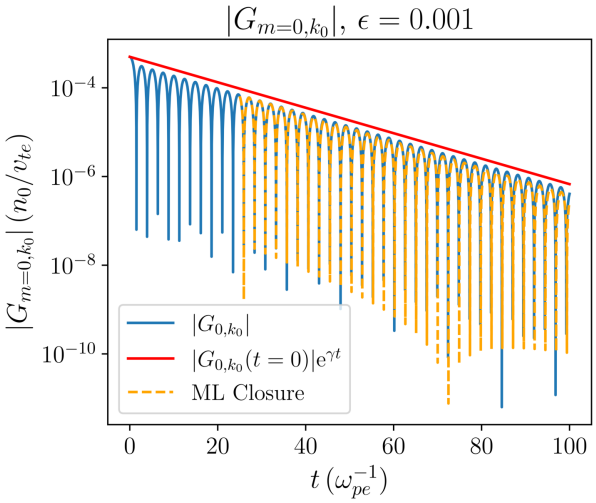



Figure 4.4: Comparison between baseline numerical solution, ML closure, and theoretical damping rate of the Fourier-Hermite amplitude for an initial cosine density perturbation. The low-amplitude perturbation shows strong agreement with the theoretical damping rate. When augmented with the ML closure, the moment solver continues to capture the behavior well at a lower Hermite resolution of $M = 4$, as opposed to the $M = 17$ baseline.

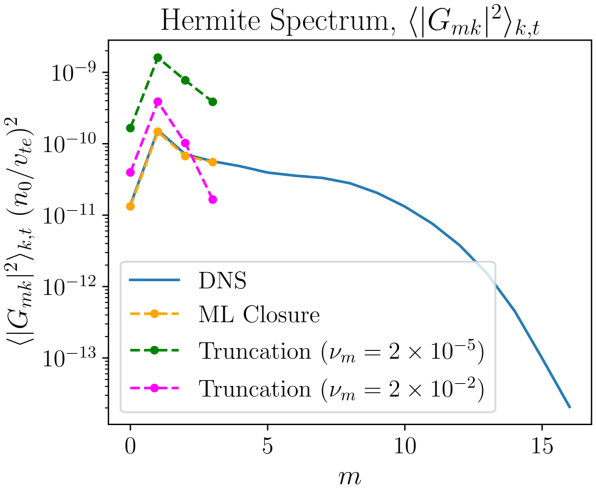



Figure 4.5: Hermite spectra of the high-velocity-resolution ($M = 17$), truncated ($M = 4$) simulations, and ML closure for the initial value problem in Figure 4.4. The spectra are averaged over time and Fourier wavenumber. The ML closure model permits a low-resolution simulation to accurately resolve the Hermite spectrum.

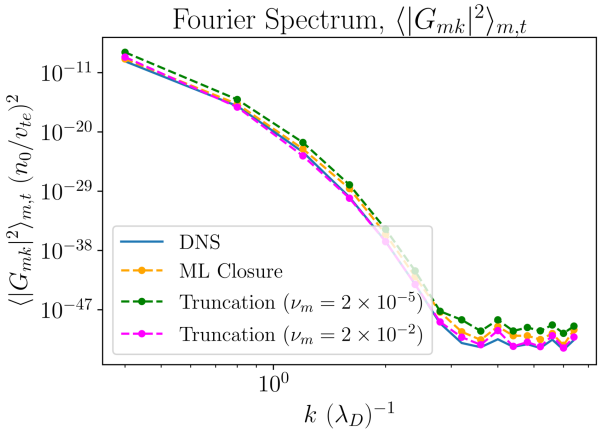



Figure 4.6: Fourier spectra of the simulations in Figure 4.5. In this weakly nonlinear regime, the majority of the energy remains in the boxscale mode. Each reduced Hermite resolution model properly captures the rapid energy decay across $k$.

flux of energy flows between wavenumbers. To create a low-resolution simulation without the ML closure, we truncate the moment hierarchy at $m_c = 3$ and set $G_{m_c+1,k} = 0$ in the evolution equation for $G_{m_c,k}$. As in Section 4.3.1, we also compare to a low-resolution simulation with increased hypercollisional regularization, with $\nu_m$ set to $2 \times 10^{-2}$ to maintain finite dissipation at the finest resolved scales. An intuitive explanation for the poor performance of the truncated simulations in Figure 4.5 can be described in relation to the general form of the evolution equation for this problem, (3.91). The truncation process removes the $G_{m+1,k}$ term from the final moment equation in the truncated system, eliminating an effective energy sink from that moment. Without access to the proper dissipation channel of smaller scales in velocity space, the truncated system experiences some energy reflection and pile up in its smallest resolved scales in velocity. In contrast, the ML closure learns the pattern of dissipation through Landau damping and serves as a better boundary condition than truncation.

We find that the reservoir hyperparameters that result in the best closure performance for this initial condition are spectral radius $\rho_{sp} = 0.6$, Tikhonov regularization parameter $\beta = 10^{-7}$, and $T = 25$ normalized time units for training. For each wavenumber, we set the reservoir input scaling parameter $\sigma$ to normalize the reservoir inputs by the time average of the Hermite amplitude at the moment before the closure:

$$\sigma_k = \frac{1}{\langle |G_{m_c-1,k,t}| \rangle_t}. \tag{4.14}$$

This preserves the dynamic range of the hyperbolic tangent nonlinear activation function by mitigating the possibility of saturation to $-1$ or 1. As in Section 4.3.1, we set the degree of the adjacency matrix to three. Finally, we choose a reservoir size of two nodes per input value, totaling twelve nodes.

### 4.3.3 Strongly Nonlinear Vlasov-Poisson

Finally, we create an ML closure model for the strongly nonlinear regime of the Vlasov-Poisson system. When the amplitude of the initial density perturbation becomes significant relative to the background, the nonlinear term dominates over the energy cascade in linear Landau damping. Figure 4.7 presents the numerical solution to the initial value problem from Section 4.3.2, with $\epsilon$ increased to 0.18. When compared to Figure 4.4, the behavior at high amplitude deviates from the linear theory. Additionally, a convergence study indicated that more Hermite moments ($M = 65$) are necessary to accurately resolve the low-order spectrum than in the weakly nonlinear case. (For more detail on this, see Appendix A.1.) We set $\nu_m$ to $5 \times 10^{-7}$, following the procedure we used to determine the baseline hypercollisional regularization parameter for the linearized and weakly nonlinear cases. From the start of the simulation until approximately $t = 25$, the mode damps at a faster rate than in the linear theory. After the initial transient, the mode saturates to a slower damping rate than Landau's prediction. When the ML closure model is introduced, the lower Hermite resolution ($M = 4$) system maintains an accurate frequency and slightly overdamps the amplitude. The training phase for the ML closure model was set to $50 \leq t < 100$ to avoid the most severe initial transient behavior, and the time trace for the ML closure begins at $t = 100$.

We can quantitatively determine whether particle trapping arrests the damping rate in Figure 4.7 by calculating the particle bounce time for its initial condition and comparing it to the low-frequency periodic oscillations in the simulation. In his seminal work, O'Neil [1965] derived an analytic expression for the damping rate as a function of time for collisionless nonlinear Landau damping with a sinusoidal electric field. He argued that the physical interpretation for the decay of the damping rate is that trapped electrons undergo cyclic oscillations in their density on a non-

linear timescale which is slower than the characteristic timescale of linear Landau damping. This phenomenon has been observed in nonlinear beam plasma experiments [Dimonte and Malmberg, 1978].

We can follow O'Neil [1965] to calculate the characteristic bounce time for the simulation. In a frame of reference that moves with a propagating plasma wave, conservation of energy for a single electron takes the form

$$W = \frac{1}{2}mv^2 - e\Phi_0 \cos(kz), \tag{4.15}$$

where $\Phi_0$ is the initial amplitude of electrostatic potential. Solving this equation for velocity yields

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \pm\sqrt{\frac{2}{m}[(W + e\Phi_0 \cos(kz)]}. \tag{4.16}$$

Electrons become trapped when they do not have enough energy to escape the potential well of the wave. Under that condition, the particles bounce and their velocities flip signs. In (4.16), velocity can reverse directions if $W < e\Phi_0$. To find the bounce time, $\tau_b$ for this scenario, one can separate the differentials in (4.16) and integrate. The result is

$$\tau_b = \frac{4K(\kappa^2)}{k}\sqrt{\frac{m}{e\Phi_0}}, \tag{4.17}$$

where $K(\kappa^2)$ is the complete elliptic integral of the first kind, and we have defined its argument as the dimensionless parameter

$$\kappa^2 \equiv \frac{1}{2}\left(\frac{W}{e\Phi_0} + 1\right). \tag{4.18}$$

Normalizing wavenumber $k$ by the Debye length $\lambda_{\mathrm{De}}$, time $t$ by the inverse plasma frequency

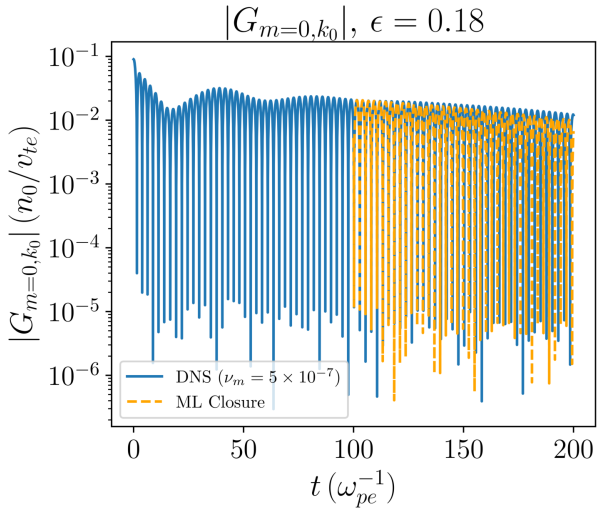



Figure 4.7: Simulated Fourier-Hermite amplitude for a high initial amplitude (18% of background) cosine density perturbation. The dynamics exhibit strongly nonlinear behavior, attenuating the damping of the mode. As in the low-amplitude case, the ML closure model captures the frequency and amplitude of the wave well.

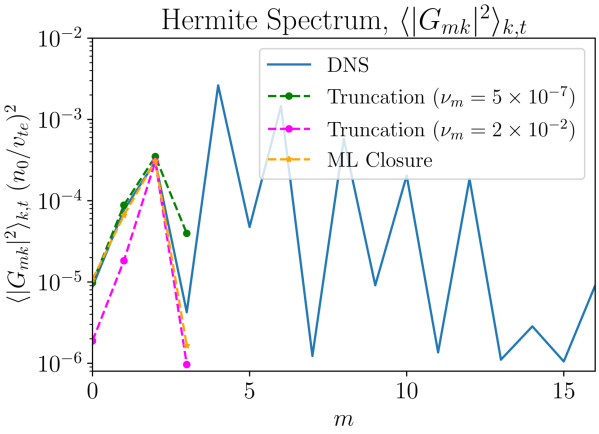



Figure 4.8: Hermite spectra of the high-velocity-resolution ($M = 65$), truncated ($M = 4$) simulations, and ML closure for the initial value problem in Figure 4.7 averaged over $k$ and $t$. While both the ML closure model and truncated simulation agree with the high-resolution direct numerical simulation, the ML closure model shows closer agreement.

$\omega_p e^{-1}$, mass $m$ by the electron mass $m_e$, and charge by the elementary charge results in the condition

$$\tau_b = \frac{4K(\kappa^2)}{k} \frac{1}{\sqrt{\Phi_0}}. \tag{4.19}$$

Using an estimate of $K \approx 2$ for most trapped particles for the $\Phi_0 = 0.18$ and $k = 0.4$ case yields $\tau_b \approx 47$. The peaks in the low-frequency portion of the nonlinear oscillation in Figure 4.7 are spaced approximately 45 $t\omega_{pe}$ apart. The slow damping rate and low-frequency oscillations in Figure 4.7 are consistent with the theory of nonlinear trapping of electrons.

As in the previous section, Figure 4.8 displays a comparison between the Hermite spectra of the high-resolution baseline, the ML closure model, and two truncated low-resolution simulations,

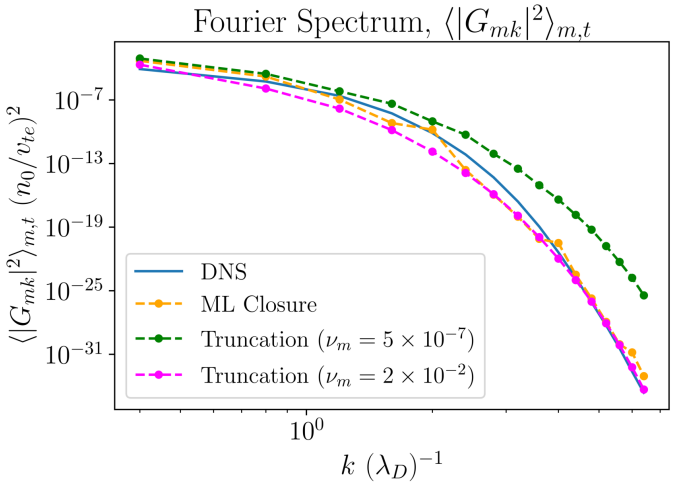



Figure 4.9: Fourier spectra of the simulations in Figure 4.8 . The ML closure resolves the wavenumber spectrum more accurately than the truncated simulations, improving simulation accuracy at low resolution in velocity space.

where one has an increased amount of hypercollisional regularization to account for the lower resolution. At this higher initial amplitude, the Hermite spectrum of the baseline does not decrease monotonically, creating a more challenging scenario for the reservoirs to model. At the same value of $\nu_m$, both the truncated low-resolution simulation and the ML closure model accurately capture the Hermite spectrum of the system at low $m$, but as Figure 4.9 reveals, only the ML closure model also agrees with the Fourier spectrum. Mean-squared errors in the ML closure spectra are $3.66 \times 10^{-18}$ and $1.19 \times 10^{-14}$ for the Hermite and Fourier spectra, respectively. The ML closure successfully captures both spectra of the system, reducing the required velocity-space resolution by a factor of sixteen. We observe some oscillatory behavior in the Fourier spectrum of the ML closure, but our tests do not identify a mechanism that causes this behavior. We obtain the reservoir hyperparameters by the same method as the previous section, adjusting them to $\beta = 10^{-6}$, $T = 50$, and 120 nodes in each reservoir. Other parameters remain the same.

A potentially surprising result is that the Hermite spectrum of the naïvely truncated sim-

ulation is more accurate than its Fourier spectrum, despite the fact that the truncation occurs in Hermite space. We intuit that this result is due to a pile up of excess free energy at the closure moment, which then nonlinearly advects in Fourier space. Adjusting the strength of the hyper-collisions slightly improves the accuracy of the Fourier spectrum, at the cost of overdamping the density and momentum moments. An analysis of the flux of free energy in this system similar to the work of Meyrand et al. [2019] may be used to investigate this result further, but that is beyond the scope of this work.

## 4.4 Summary and Conclusion

In this chapter, we have used reservoir computing to present an ML, velocity-space closure model for the hierarchy of Hermite moments in the one-dimensional Vlasov-Poisson system. The closure model serves as a proof-of-principle that reservoir computing can be used to reduce the simulation domain requirements for studying plasma dynamics.

A crucial extension of this work is to train the models to generalize to different initial conditions or forcing models. Gyrokinetic turbulence simulations are expensive, so we wish to reuse as much training data as possible in pursuit of the most powerful models. Moreover, we wish to avoid the need to run high-fidelity simulations before switching over to the closure model. Although training reservoir computing models is inexpensive, using pretrained closure models will be faster than training them "on-the-fly."

A logical objective to pursue is to extend the closure to reduce the required resolution in the Fourier representation of position space. In practice, gyrokinetic turbulence simulations commonly use significantly more resolution for configuration space than they do for velocity space, in part

because of the near-Maxwellian nature of the velocity distributions. Therefore, a Fourier-space closure model would potentially be more desirable than a velocity-space closure. However, the structure of the nonlinear term poses a challenge to achieving that goal. In the Vlasov-Poisson system, the coupling between resolved and unresolved scales in velocity space occurs as a sum of nearest-neighbors in Hermite space. This locality in the spectral domain defines a clear, single closure term. In contrast, the convolution in the nonlinear term establishes nonlocal interactions across scales in Fourier space. An ML closure for Fourier space would require a different structure than the one developed for this chapter.

The eventual goal is to build closure models that can reduce the domain requirements for higher dimensional turbulence simulations, including the full gyrokinetic equation. A next step toward that goal would be to test this velocity-space closure in that system. There has been recent interest in applying spectral formulations of velocity space to gyrokinetic simulations, particularly with a Hermite basis used for parallel velocity and a Laguerre polynomial basis for perpendicular velocity [Jorge et al., 2017; Mandell et al., 2018; Hoffmann et al., 2023b,a; Frei et al., 2023, 2024]. A successful implementation of an ML closure for Fourier-Hermite-Laguerre codes, including `Gyacomo` [Hoffmann et al., 2023a] and `GX` [Mandell et al., 2024] may improve their capabilities to resolve turbulence statistics with lower resolution requirements. Empirically, both codes calculate turbulent heat fluxes with high accuracy using eight or fewer Laguerre modes, suggesting that future work on ML closures should prioritize reductions in spatial and parallel velocity resolution requirements.

An ideal closure model would have a compact, symbolic form from which physical intuition can be derived. Significant recent progress toward interpretable, symbolic closures has been achieved using sparse regression techniques [Donaghy and Germaschewski, 2023; Cheng et al.,

2023; Ingelsten et al., 2024]. One challenge that these algorithms face is the difficulty of finding symbolic closures that are not local in space. The analytic closure derived by Hammett and Perkins [1990] for the linear limit of the system in this chapter is local in Fourier space, but it includes a nonlocal Hilbert transform when expressed in configuration space. Closures for the nonlinear form of this system or other systems of interest may also require information that is nonlocal in configuration space, but local in a spectral representation. The ML closure developed in this chapter and by Huang et al. [2025] apply a Fourier representation to the data, providing the neural networks direct access to locality in spatial scales, and the velocity-space closure in this chapter is also local in Hermite space. Though interpreting moment closures that use neural networks is more challenging than closures derived by sparse regression techniques, future work may incorporate feature importance identification methods like those formalized by Lundberg and Lee [2017].

# Chapter 5: Progress Toward Automatic Identification of Under-resolved Gyrokinetic Turbulence Simulations: Data-Driven Heuristics

As a brief summary, in Chapter 3, we discussed that there are a myriad of factors to consider when designing a numerical scheme to study dynamics in turbulent systems, and we motivated our choice of an Eulerian pseudo-spectral method to study our problems of interest. However, once an approach to solving the PDE has been chosen, questions remain about what the physical domain should be and how much resolution should be used to represent it. In Chapter 4, we introduced an ML closure method that decreases resolution requirements in velocity space for the 1-D Vlasov-Poisson system and discussed some possible extensions to higher-dimensional systems like the 5-D electrostatic gyrokinetic system. In this chapter, we seek to develop some heuristics that may provide insights into these lingering questions.

Machine learning or other data-driven models rely on the existence of reliable and robust datasets on which these models can be trained. Indeed, in recent years, improvements in memory, bandwidth, and processing speed, in addition to the paradigm shift toward parallel processing on accelerator platforms like GPUs, have facilitated gyrokinetic simulations at scale. Some examples of this include the optimization of stellarator magnetic geometries to reduce turbulent heat flux [Kim et al., 2024] and the development of an $\mathcal{O}[10^5]$ element database of gyrokinetic simulations to create an ML surrogate model for the mapping between flux tube geometry and heat flux due to ITG turbulence [Landreman et al., 2025]. While it is desirable to accelerate these simulations, it is also imperative to ensure that the simulation results that underpin the training of a reduced model are valid.

With the advent of workflows that involve the autonomous execution of thousands of turbulence simulations, tedious manual inspection of simulation results to ensure accuracy becomes infeasible. Simliarly, full convergence studies requiring ensembles of simulations in a structured scan over domain input parameters may be cost-prohibitive. As an example, consider a two-parameter scan over resolution and domain extent in the perpendicular plane, maintaining a fixed ratio between both coordinates perpendicular to the magnetic field. A search to find the optimal combination of these input parameters using a coarse grid could require several hours, as convergence to statistical steady state typically requires between two and forty minutes of wall clock time, depending on the number of points and spectral moments used to define the simulation grid. A valuable objective would be to develop some heuristics or metrics that would empower users to conduct large-scale parameter scans or studies without the need to manually confirm the validity of the simulation or to perform these costly convergence studies for each case.

In this chapter, we investigate the potential use of several candidate diagnostics to identify when the domain and resolution of a gyrokinetic turbulence simulation does not properly resolve. To assess the validity of simulation domain, we consider several numerical reductions of correlation functions of the turbulence and assess their impacts on the turbulent heat flux. We also study the potential of a nonlinear energy transfer diagnostic to indicate a lack of sufficient resolution in simulations. Finally, we observe turbulent heat flux and free energy spectra for markers of insufficient resolution.

The long-term goal of this work is to integrate this feature into `GX` as automated verification of domain and resolution inputs. As a first step toward that goal, we implement these diagnostics as post-processing tool in Python. Toward the end of this chapter, we discuss strategies for future work to implement these diagnostics and the heuristics derived from them *in-situ* as CUDA/HIP

kernels for concurrent execution on the GPU(s) during a simulation.

## 5.1 Simulation details

In this chapter, we use GX to study two different magnetic geometries, one tokamak and one stellarator. For the tokamak case, we study the "Cyclone Base Case" [Dimits et al., 2000], a commonly used ITG benchmark scenario based on the Doublet III-D (DIII-D) tokamak run by General Atomics in San Diego, California. For the stellarator case, we choose the GX stellarator benchmark case, which is based on Wendelstein 7-X (W7-X), an experimental stellarator operating at the Max Planck Institute for Plasma Physics in Greifswald, Germany.

The domain and resolution parameters we vary in this study are the resolution in the perpendicular plane $N_x$ and $N_y$ and the domain extent in the perpendicular plane $L_y$. We use a flux tube that is approximately square at the outboard midplane such that $L_x \approx L_y$. Due to the twist-and-shift parallel boundary condition derived by Beer et al. [1995], $L_y$ and $L_x$ are related by the magnetic shear and are not completely independent parameters. The physical periodicity of the system lies in the toroidal angle, $\zeta$, and the poloidal angle, $\theta$. In systems with magnetic shear, the ratio between $L_x$ and $L_y$ is quantized such that the quantity

$$\delta j = 2\pi N \hat{s} L_x / L_y \tag{5.1}$$

is an integer, where $\hat{s}$ is the magnetic shear of the flux tube, and $N$ is an arbitrary integer. Thus, for given values of $\hat{s}$ and $L_y$, we choose $\delta j$ to be the closest integer to $2\pi\hat{s}$, so that we can obtain a roughly square box at the outboard midplane. The physics parameters we vary are the temperature gradient scale length and the ratio between the electron temperature and ion temperature. As de-

scribed in Chapter 2, these simulations are electrostatic ion temperature gradient (ITG) turbulence simulations with adiabatic electrons, and we use only one ion species.

For the Cyclone Base Case, we use a Miller parameterization [Miller et al., 1998] to represent the magnetic equilibrium. The Miller parameterization is a local, analytic solution to the Grad-Shafranov equation [Grad and Rubin, 1958] for MHD force balance, where in cylindrical $(R, Z)$ coordinates, the flux surface is specified by

$$R(r, \theta) = R_0(r) + r\cos[\theta + (\sin\theta)(\arcsin\delta(r))], \tag{5.2}$$

$$Z(r, \theta) = \kappa(r) r \sin\theta, \tag{5.3}$$

where $\theta$ is the poloidal angle, $R_0$ is major radius at the magnetic axis, $r$ is the half-diameter of the flux surface, $\kappa$ is the elongation, and $\delta$ is the triangularity. We specify a circular flux surface and local equilibrium with the following geometric parameters: $R_0/a = 2.77778$, $r/a = 0.5$, elongation $\kappa = 1$, and triangularity $\delta = 0$, and the gradients of the elongation and triangularity are zero. The left side of Figure 5.1 shows the flux surface we use for the cyclone base case. The global magnetic shear is $\hat{s} = 0.8$, and the safety factor is $q = 1.4$. We choose a normalized density gradient of $a/L_n = 0.8$, and we perform a scan over normalized ion temperature gradients such that $a/L_T \in \{2.5, 3.5, 5, 10\}$. For the cyclone base case simulations, we use $N_\theta = 24$ equally spaced grid points in $\theta$ and coarse velocity space resolution with $N_m = 8$ Hermite and $N_\ell = 4$ Laguerre moments for $v_\parallel$ and $\mu B$ respectively. A prior velocity convergence study for the cyclone base case performed by Mandell et al. [2024] showed that this modest amount of velocity space resolution is sufficient for the problem. Each simulation spans 800 $a/v_t$ normalized time units. For

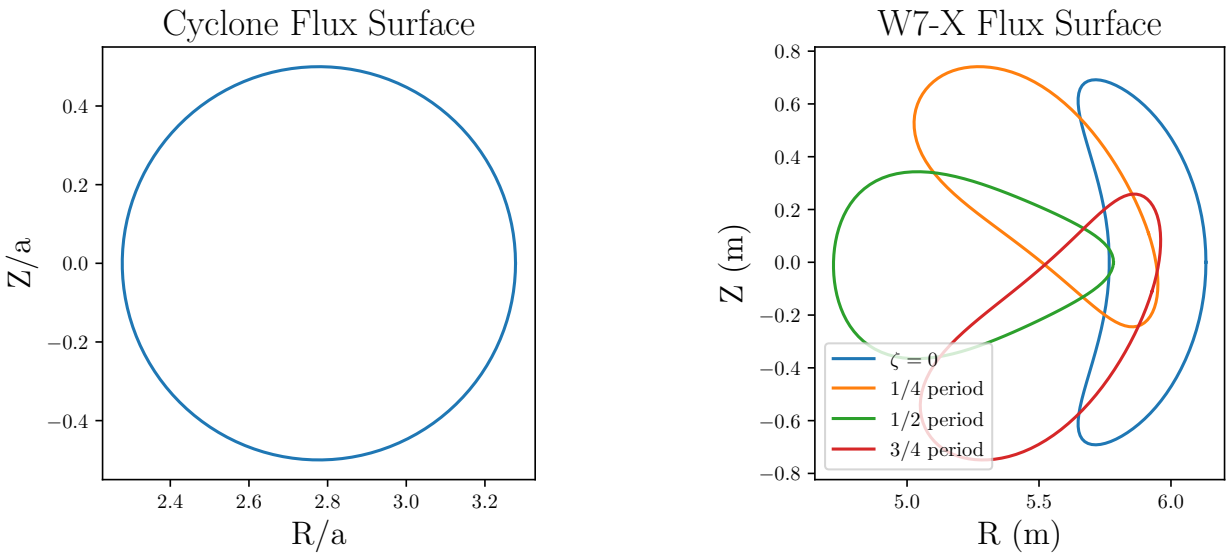



Figure 5.1: Visualizations of the flux surfaces for the ITG simulations in this thesis in cylindrical $(R, Z)$ coordinates. The cyclone base case uses a circular flux surface (left) constructed using a local Miller parameterization [Miller et al., 1998]. For W7-X, we choose the $s = 0.64$ flux surface from a `VMEC` equilibrium (right). Poloidal cross sections are shown at four equally spaced locations in toroidal angle $\zeta$ for one field period.

context, with the DIII-D minor radius of 0.6 m and an ion temperature of 1 keV, this simulation duration would correspond to roughly 3 $\mu$s for a deuterium plasma.

For the W7-X simulations, we use the `GX` stellarator benchmark case, which is a flux tube chosen from an MHD equilibrium calculated by the `VMEC` code [Hirshman and Whitson, 1983] and was used to benchmark `GX`, `stella`, and `GENE` [González-Jerez et al., 2022; Mandell et al., 2024]. The W7-X simulations require more resolution along the field line to resolve the non-axisymmetric geometry, with $N_\theta = 48$. We again use $N_m = 8$ and $N_\ell = 4$ for the velocity resolution and allow the simulation to run for 800 $a/v_t$ time units to reach a statistical steady state.

## 5.2 Correlation Functions

In this section, we explore the potential use of correlation functions of the turbulence to diagnose whether the simulation domain is improperly sized. The gyrokinetic ordering assumptions formalize the anisotropy of spatial scales, so we expect to find a qualitative difference between parallel and perpendicular correlation length requirements.

### 5.2.1 Parallel Correlations: Critical Balance

Indeed, analyses of parallel correlations have been used to study scaling laws in ITG turbulence, relating parallel and perpendicular length scales to each other [Barnes et al., 2011; Nies et al., 2026]. To assess their theory of critically balanced turbulence, Barnes et al. [2011] compared their scaling law prediction,

$$\frac{\ell_\parallel}{qR} \sim \left(\frac{\ell_\perp}{\rho_i}\frac{L_T}{qR}\right)^{4/3}, \tag{5.4}$$

to parallel correlation statistics calculated from `GS2` [Dorland et al., 2000] simulations of the cyclone base case. They define a parallel correlation function

$$\mathcal{C}(k_y, \Delta\theta) = \frac{\sum_{k_x} \Phi(k_x, k_y, \theta = 0)\Phi^*(k_x, k_y, \theta = \Delta\theta)}{\sum_{k_x} |\Phi(k_x, k_y, \theta = 0)|^2}, \tag{5.5}$$

where subscripts $x$ and $y$ correspond to the magnetic flux surface and field line labels respectively, and $\theta$ is the poloidal angle. Integrating the correlation function yields a parallel correlation length:

$$\overline{\Delta\theta}(k_y) \equiv \int \mathrm{d}(\Delta\theta)\mathcal{C}(k_y, \Delta\theta). \tag{5.6}$$

This metric can be used as a measure of the parallel length scale in (5.4).

In this work, we implement correlation length diagnostics for the GX gyrokinetic code Mandell et al. [2024]. For a first example, we consider the same case as the original ITG critical balance paper [Barnes et al., 2011]. This simulation uses a spatial resolution of $(N_x, N_y, N_\theta) = (192, 192, 32)$, and the perpendicular box size $L_y$ is set to $L_y = 200\pi\rho_i$ such that the smallest resolved binormal wavenumber is $k_y\rho_i = 0.1$. Figure 5.2 shows the time-averaged parallel correlation length (5.6) as a function of $k_y$ for this simulation, with a dashed line plotting the critical balance scaling. Figure 5.3 shows the same diagnostic, plotted for the W7-X reference case, with $(N_x, N_y, N_\theta) = (96, 96, 48)$ and $L_y = 42\pi\rho_i$. The sharp peaking in the correlation function at $\theta = 0$ suggests that the turbulence rapidly decorrelates along the field line.

In comparison to the cyclone base case, the W-7X case shows far weaker correlations parallel to the magnetic field. One plausible explanation is the presence of magnetic field period symmetries in the stellarator case. As explained in Section 5.1, the cyclone base case uses a circular flux surface from an axisymmetric tokamak. In contrast, the W7-X equilibrium magnetic field has a five-field-period symmetry [Geiger et al., 2015], meaning that for poloidal and toroidal angles $(\theta, \zeta)$,

$$\mathbf{B}_{\mathrm{W7-X}}(\theta, \zeta + 2\pi/5) = \mathbf{B}_{\mathrm{W7-X}}(\theta, \zeta). \tag{5.7}$$

Although this symmetry is in the toroidal angle, the field period symmetry could set an effective cap on parallel correlation lengths in $\theta$ that could depend on the rotational transform. This could partially explain the discrepancy between the critical balance scaling prediction and the short parallel correlation lengths in Figure 5.3. A study focused on evaluating parallel correlation functions for a range of values of rotational transform and field period symmetries may provide the data

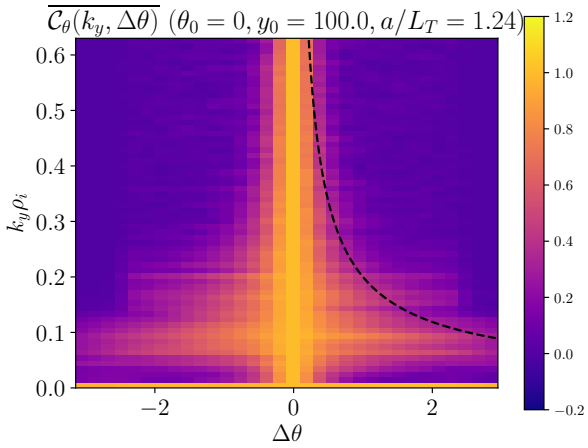



Figure 5.2: Parallel correlation function for a GX simulation of the cyclone base case using parameters similar to those in [Barnes et al., 2019]. The dashed line corresponds to the critical balance scaling law $(\ell_\parallel/R) \sim (\ell_\perp/\rho_i)^{4/3}$, where we take $\overline{\Delta\theta}$ as a measure of $\ell_\parallel/R$ and $k_y\rho_i$ as a measure of $\ell_\perp/\rho_i$.

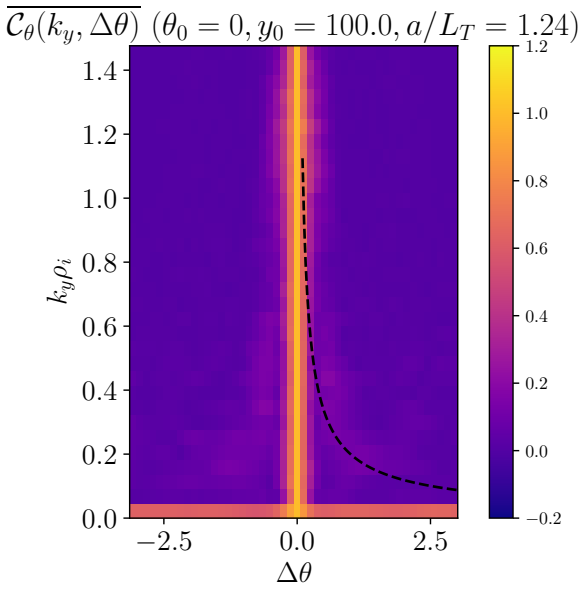



Figure 5.3: Parallel correlation function for an ITG turbulence simulation of the $\alpha = 0$ field line on the flux surface in the right of Figure 5.1. As in Figure 5.2, the dotted line corresponds to the crtical balance scaling. The stellarator case indicates that the turbulence shows weak correlation in $\theta$.

necessary for a quantitative assessment of this explanation.

By the gyrokinetic orderings, the parallel dynamics are macroscopic, while the perpendicular dynamics occur at scales on the order of the ion gyroradius. Details of the magnetic geometry which enter into the gyrokinetic equation, including $|B|$, $\nabla x$, and $\nabla y$, vary along a flux tube at scales comparable to the normalizing system scale $a$. In this work, we take $a$ to be the minor radius, but other choices like the major radius are also suitable. In our simulations, we find that parallel correlations are too short to form a useful minimum flux tube length. For this reason, we devote the rest of our study to perpendicular domain and resolution specifications.

### 5.2.2 Perpendicular Correlations

#### 5.2.2.1 Building Perpendicular Correlation Functions for the Cyclone Base Case

It is possible to create multiple perpendicular correlation functions for the electrostatic potential. To build some intuition for the dynamics and to make a more informed decision about which perpendicular correlation function to use for a heuristic, we will begin by examining data from the cyclone base case. Figure 5.4 presents the first sample we consider, the electric potential $\Phi$, evaluated as a function of $x$ and $y$ at the outboard midplane $\theta = 0$, 1,000 (left) and 5,000 (right) timesteps after initialization of an ITG simulation with $Nx = 196$, $Ny = 64$, $N_\theta = 24$, and $L_y = 56\pi$. The heat flux time trace for this example is shown in Figure 5.5. Though the code uses a Fourier spatial representation, we apply the inverse Fourier transform to the electrostatic potential to evaluate it on the real-space grid. This facilitates a more straightforward interpretation of its features in relation to its correlation length.

At these timesteps, the electric potential at the outboard midplane appears to exhibit periodic

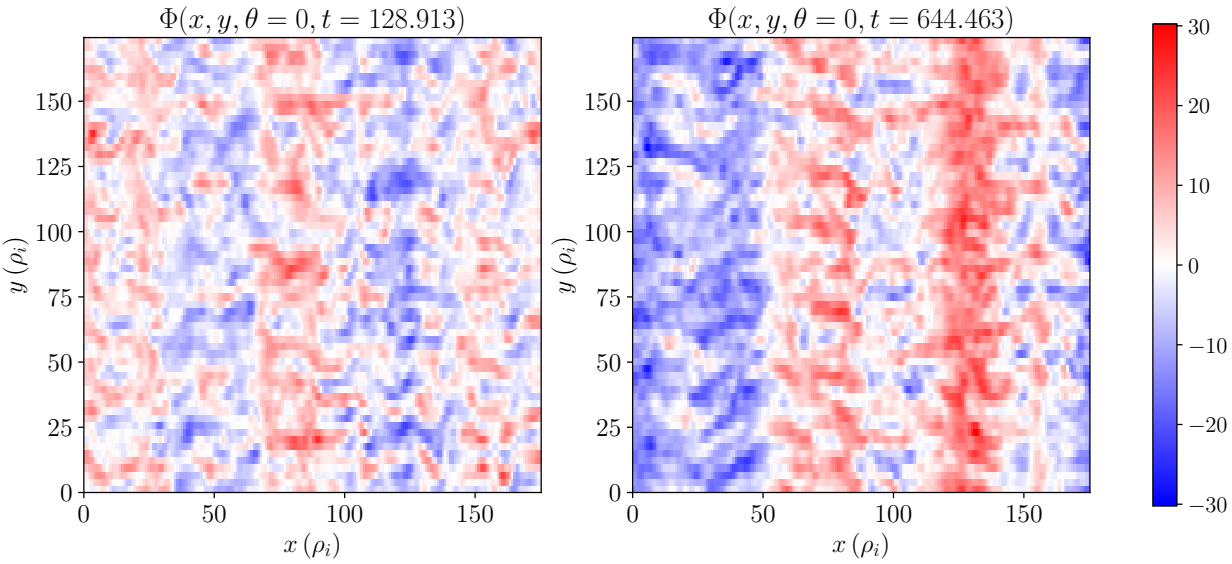



Figure 5.4: Electric potential in the cyclone base case, evaluated at the outboard midplane 1,000 timesteps (left) and 5,000 timesteps (right) after initialization.

oscillations in both the $x$ and $y$ directions, at wavelengths smaller than the box scale at which we impose periodicity. The cyclone base case is known to be a zonal-flow dominated scenario, so the apparent zonal mode with $k_y = 0$ is expected. We see strong evidence of it around $x = 125\rho_i$. We examine this feature in more detail in Figure 5.6 by plotting $\Phi(y)$ at the radial position $x = 124.8\rho_i$. This clearly demonstrates the strong zonal mode, as the potential oscillates around a mean of $-9.82$.

#### 5.2.2.2 Averages of the Correlation Function

Instantaneous, point-wise correlation functions can have spurious correlations that are not representative of the typical dynamics of the system. We can instead take averages of the correlation function to avoid this issue. We will start with the $x$ correlations, where the impact of the strong zonal flow in $y$ will be less prominent. To investigate the spatial homogeneity of the $x$

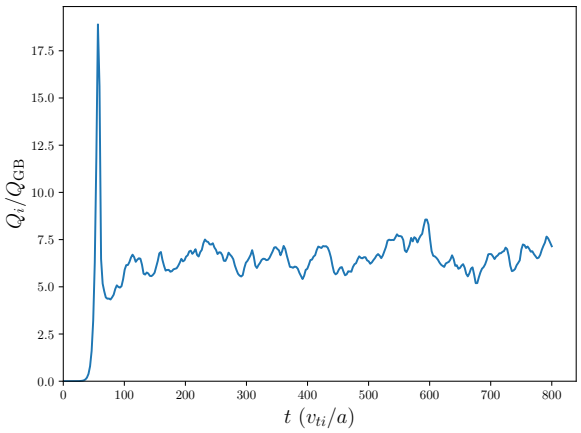



Figure 5.5: Heat flux time trace for the cyclone base case. The initial transient ends by $t(v_{ti}/a) = 100$.

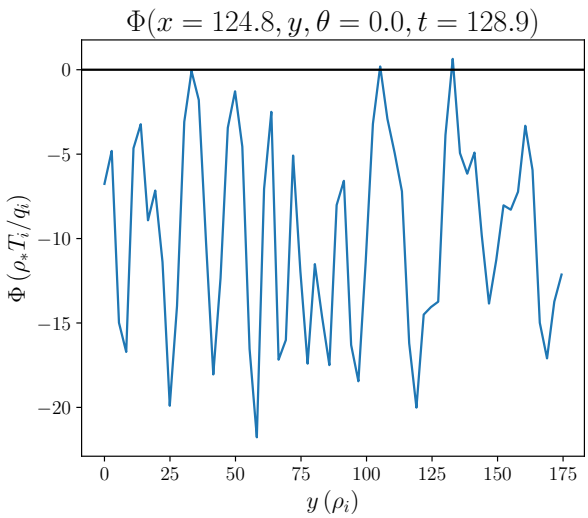



Figure 5.6: Electric potential from the left plot of Figure 5.4, evaluated as a function of $y$ at $x = 124.8$.

correlations, we can compute their averages in $x$ and

$$\langle \mathcal{C}_x \rangle_x (y, \theta, t, \Delta x) = \frac{\frac{1}{N_x} \sum_x \left( \Phi(x + \Delta x, y, \theta, t) \Phi(x, y, \theta, t) \right)}{\frac{1}{N_x} \sum_x \left( \Phi(x, y, \theta, t) \right)^2}, \tag{5.8}$$

in $y$,

$$\langle \mathcal{C}_x \rangle_y (x, \theta, t, \Delta x) = \frac{\frac{1}{N_y} \sum_y \left( \Phi(x + \Delta x, y, \theta, t) \Phi(x, y, \theta, t) \right)}{\frac{1}{N_y} \sum_y \left( \Phi(x, y, \theta, t) \right)^2}. \tag{5.9}$$

The notation convention we use in this section is that a correlation function of the form $\langle \mathcal{C}_a \rangle_b$ measures correlations in coordinate $a$ when averaged over coordinate $b$. Figures 5.7 and 5.8 show $\langle \mathcal{C}_x \rangle_x$ and $\langle \mathcal{C}_x \rangle_y$, respectively. The diagonal banding in Figure 5.8 indicates the presence of a persistent wave.

Next, we introduce time-averaging, with the expectation that this will focus the peak of the

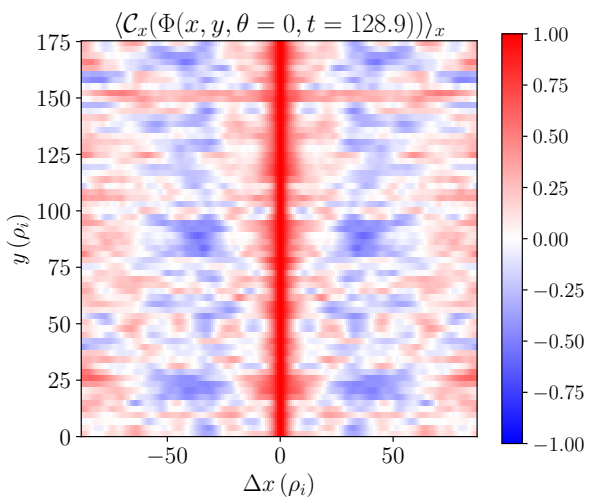



Figure 5.7: $x$-averaged $x$ correlation function for the potential in Figure 5.4 .

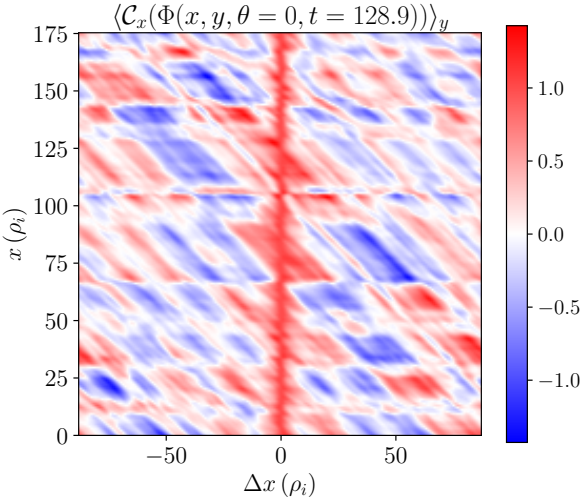



Figure 5.8: $y$-averaged $x$ correlation function for the potential in Figure 5.4 .

correlation function at $\Delta x = 0$:

$$\langle \overline{\mathcal{C}_x} \rangle_x (y, \theta, \Delta x) = \frac{\frac{1}{N_x N_t} \sum_{x,t} \left( \Phi(x + \Delta x, y, \theta, t) \Phi(x, y, \theta, t) \right)}{\frac{1}{N_x N_t} \sum_{x,t} \left( \Phi(x, y, \theta, t) \right)^2}, \tag{5.10}$$

with a similarly defined average in $y$ and $t$:

$$\langle \overline{\mathcal{C}_x} \rangle_y (x, \theta, \Delta x) = \frac{\frac{1}{N_y N_t} \sum_{y,t} \left( \Phi(x + \Delta x, y, \theta, t) \Phi(x, y, \theta, t) \right)}{\frac{1}{N_y N_t} \sum_{y,t} \left( \Phi(x, y, \theta, t) \right)^2}. \tag{5.11}$$

Time averages are represented by an overline and are calculated using data after the heat flux time traces reach a statistical steady state. Figures 5.9 and 5.10 show the $x, t$ and $y, t$ averages of the $x$ correlations. As expected, a prominent vertical band at $\Delta x = 0$ emerges, but correlations at large displacements remain significant. This is due to the $k_y = 0$ zonal flow, which is intuitively noticeable in Figure 5.9.

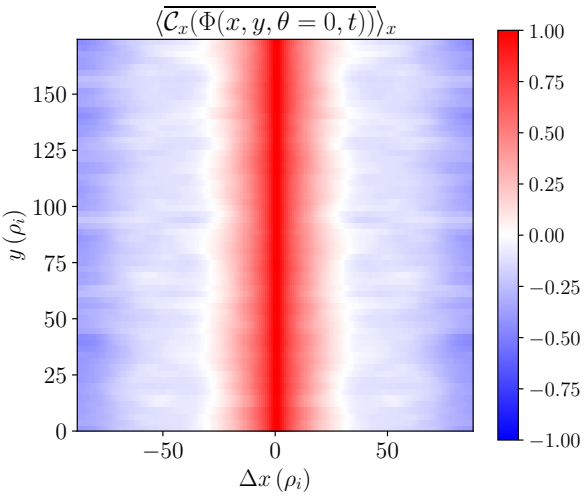



Figure 5.9: Correlation function in $x$ averaged over $x$ and $t$ for the cyclone base case.

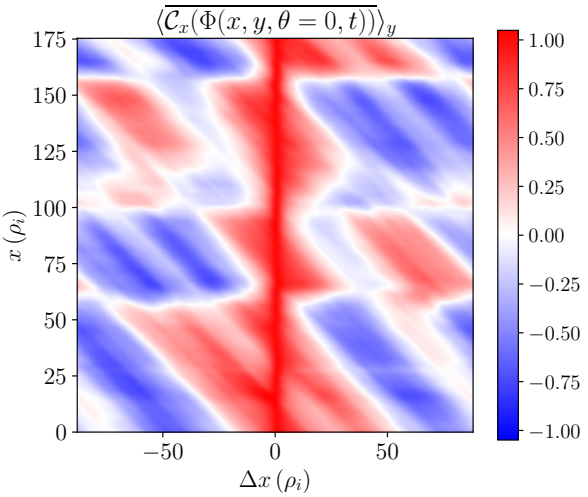



Figure 5.10: Correlation function in $x$ averaged over $y$ and $t$ for the cyclone base case.

The zonal mode structures in the cyclone base case will require more careful analysis. We anticipate that the fluctuations may be homogeneous in $x$ and $y$, but the zonal modes create an effectively infinite-wavelength correlation in $y$. A better approach might be to measure correlations in the fluctuations in $y$ after subtracting the mean. To illustrate this point, we can now plot the $y$ correlations averaged over $(x,t)$,

$$\langle\overline{\mathcal{C}_y}\rangle_x(x,\theta,\Delta y) = \frac{\frac{1}{N_x N_t}\sum_{x,t}\left(\Phi(x,y+\Delta y,\theta,t)\Phi(x,y,\theta,t)\right)}{\frac{1}{N_x N_t}\sum_{x,t}\left(\Phi(x,y,\theta,t)\right)^2}, \tag{5.12}$$

and $(y,t)$,

$$\langle\overline{\mathcal{C}_y}\rangle_y(y,\theta,\Delta y) = \frac{\frac{1}{N_y N_t}\sum_{x,t}\left(\Phi(x,y+\Delta y,\theta,t)\Phi(x,y,\theta,t)\right)}{\frac{1}{N_y N_t}\sum_{x,t}\left(\Phi(x,y,\theta,t)\right)^2}, \tag{5.13}$$

in Figures 5.11 and 5.12. These figures show near maximal correlation at all displacements in $y$. This is consistent with the presence of a zonal flow, as the $k_y = 0$ mode corresponds to an effectively infinite correlation length in $y$. Figures 5.13 and 5.14 show that removing the zonal

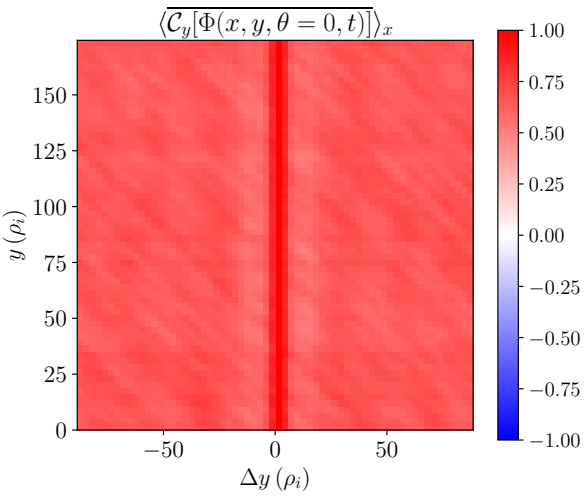



Figure 5.11: Correlation function in $y$ averaged over $x$ and $t$ for the cyclone base case. This is a clear indication of zonal-flow-dominated turbulence, as there is a strong correlation in $y$ at all displacements.

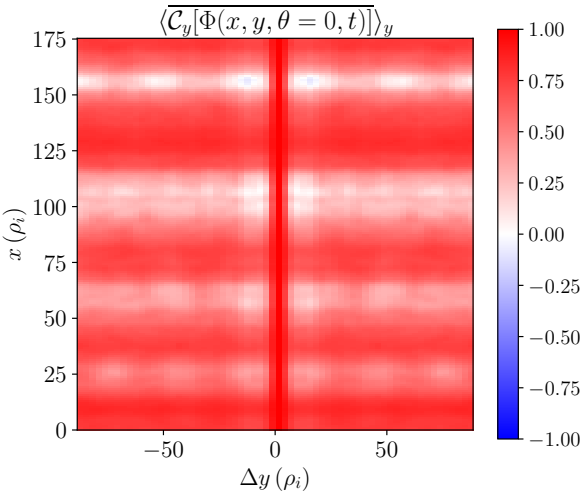



Figure 5.12: Correlation function in $y$ averaged over $y$ and $t$ for the cyclone base case. As in Figure 5.11, there is clear evidence of a zonal flow.

component of the potential yields cleaner correlation functions where the fluctuations decorrelate at small displacements from a given reference point in both $x$ and $y$.

We can finally perform spatiotemporal temporal averages over $x, y$, and $t$ simultaneously to obtain a quick diagnostic that can indicate whether fluctuations in the potential, and by extension the density, are correlated across the domain. Figures 5.15 and 5.16 show the full spatiotemporally averaged correlation functions:

$$\langle\overline{\mathcal{C}_x}\rangle_{xy}(\theta,\Delta x) = \frac{\frac{1}{N_xN_yN_t}\sum_{x,y,t}\left(\Phi(x+\Delta x,y,\theta,t)\Phi(x,y,\theta,t)\right)}{\frac{1}{N_xN_yN_t}\sum_{x,y,t}\left(\Phi(x,y,\theta,t)\right)^2} \tag{5.14}$$

$$\langle\overline{\mathcal{C}_y}\rangle_{xy}(\theta,\Delta y) = \frac{\frac{1}{N_xN_yN_t}\sum_{x,y,t}\left(\Phi(x,y+\Delta y,\theta,t)\Phi(x,y,\theta,t)\right)}{\frac{1}{N_xN_yN_t}\sum_{x,y,t}\left(\Phi(x,y,\theta,t)\right)^2}. \tag{5.15}$$

They are also plots of Figures 5.9 and 5.12, averaged over the remaining perpendicular coordi-

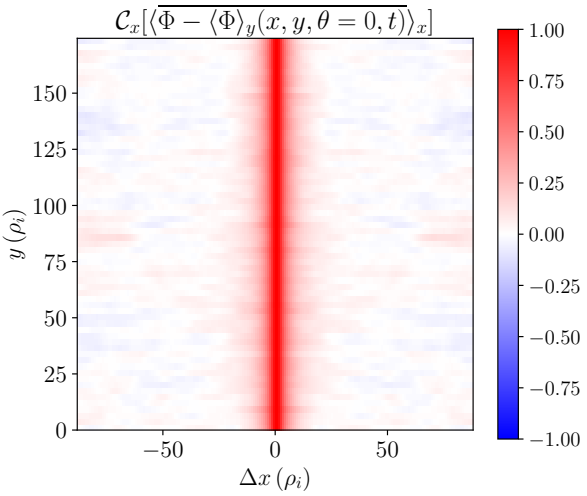



Figure 5.13: Correlation function in $x$ after removing the $ky = 0$ mode of the potential and averaging over $x$ and $t$ for the cyclone base case. A clear peak emerges around zero displacement.

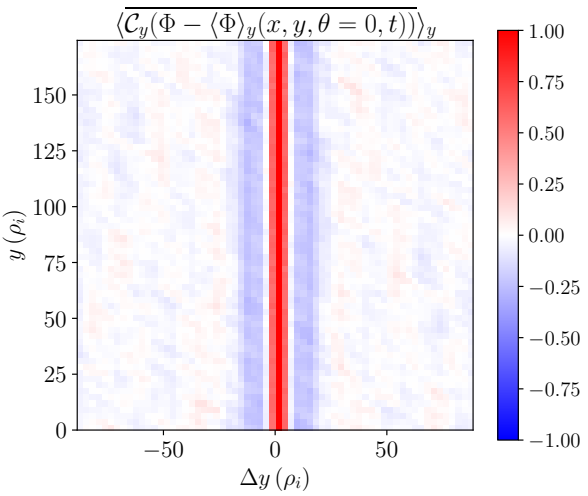



Figure 5.14: Correlation function in $y$ after removing the $ky = 0$ mode of the potential and averaging over $x$ and $t$ for the cyclone base case.

nate. The zonal flows result strong negative correlations at large displacements in $x$ due to the natural width of the zonal mode in $x$. They also produce persistent positive correlations at all displacements in $y$. These results indicate that removing the zonal mode is necessary for the $y$ correlations to be useful. The spatiotemporally averaged correlation functions of the potential align more closely with intuition once the the zonal mode has been removed. The $x$ correlations follow a decaying exponential form, while the $y$ correlations more closely resemble the product of a decaying exponential and a cosine wave. Note that we do not average over $\theta$ so that we restrict our focus to the outboard midplane where the box is approximately square. We hypothesize that for electromagnetic simulations, geometry-dependent perturbations in the magnetic field along each field line may cause $\theta$ dependence to become more important to study. That topic is left for future investigations.

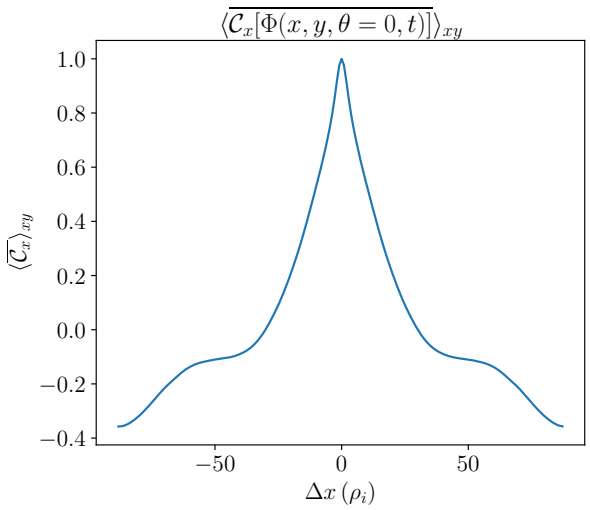



Figure 5.15: Correlation function in $x$ of the potential, averaged over $x$, $y$, and $t$ for the cyclone base case. The $k_y = 0$ zonal mode has a finite wavelength in $x$ that results in negative correlations at large displacements.

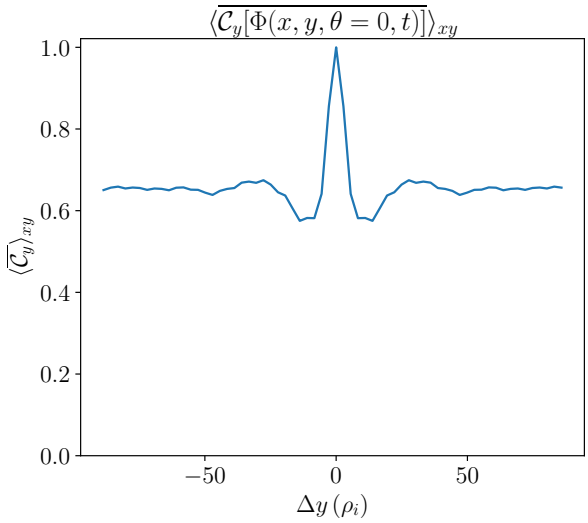



Figure 5.16: Correlation function in $y$, averaged over $x$, $y$, and $t$ for the cyclone base case. The zonal mode leads to positive displacements at all displacements.

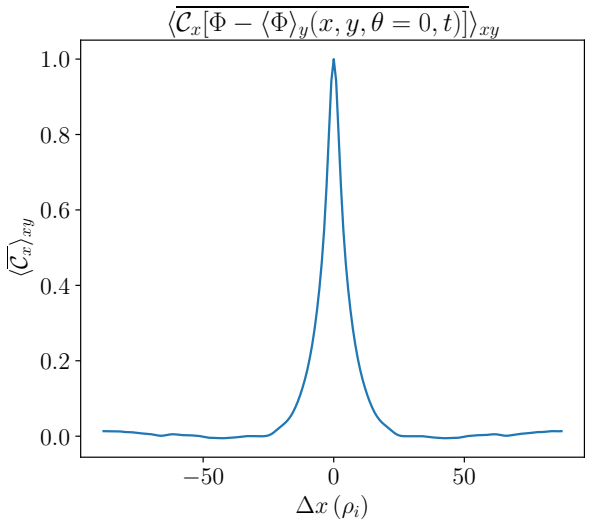



Figure 5.17: Correlation function in $x$ after removing the $ky = 0$ mode of the potential and averaging over $x$ and $t$ for the cyclone base case. As expected, the correlation of the fluctuations decays toward zero.

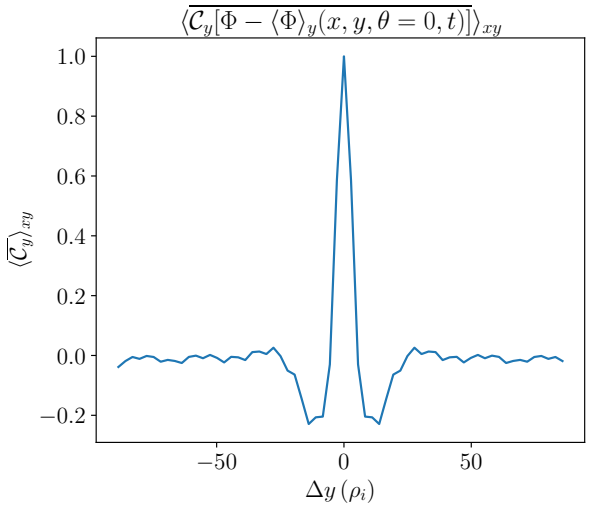



Figure 5.18: Correlation function in $y$ after removing the $ky = 0$ mode of the potential and averaging over $x$ and $t$ for the cyclone base case. Removing the zonal flow reveals oscillations in the correlation of the fluctuations, but these decay toward zero near the maximum displacements.

While the zonal flow has a formally infinite wavelength in $y$, its wavelength in $x$, $\lambda_x$ is finite. Therefore, we expect that the box size in $x$, $L_x$, should be wide enough to contain some multiples of $\lambda_x$. We posit that a set of simulations at various values of $L_x/L_y$ may yield useful data that could establish a threshold for $L_x$ in terms of the $x$ correlation length. Alternatively, a hybrid approach could be used, where we compute correlations in $x$ using the raw electric potential, and we remove the zonal flow before computing correlations in $y$.

To summarize what we have learned from this exploration of the correlation functions of this simulation of the cyclone base case:

- Single-point, single-time correlation functions of turbulent systems can introduce spurious correlations, especially when normalizing.
- Zonal flows may be removed from the signal to reveal correlations in the underlying fluctuations. Zonal flows do not have a linear instability and are known to suppress heat transport from ITG turbulence [Lin et al., 1998; Tiwari et al., 2025].
- After removing the zonal mode, correlation functions for the cyclone base case benchmark align with the intuition of statistically decorrelated turbulence across large distances in the perpendicular plane.

With this baseline understanding in hand, we will proceed to perturb the reference case in an attempt to identify a heuristic that can be used to indicate insufficient domain extent.

#### 5.2.2.3 Developing a Heuristic through a Parameter Study

In this section, we use perpendicular correlation functions to study a range of physical parameters for the cyclone base case. We also vary the box size used for the perpendicular domain to

identify the threshold at which the domain is insufficient to properly represent the physics. Because our primary objective for performing ITG simulations is to assess the impact of microtubulence on large-scale transport, we choose the heat flux $Q_i$ as our metric to assess the validity of the simulation. GX [Mandell et al., 2024] defines the radial heat flux, $Q_i$, in the electrostatic case as

$$Q_i = n_0 T_i \int \frac{\mathrm{d}x\mathrm{d}y\mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla \theta} \int \mathrm{d}^3\mathbf{v} \langle \mathbf{v}_E \rangle_{\mathbf{R}} \cdot \nabla x \left( \frac{v_\parallel^2}{2} + \mu B \right) h, \tag{5.16}$$

and the gyro-Bohm normalization is defined as $Q_{GB} = (n_0 v_t T_i)(\rho_i/a)^2$.

It would be ideal to collapse the correlation function into a single number that can be compared across different simulations. We consider correlation lengths of the turbulent fluctuations as a candidate reduction of the correlation function. In this section, we will use the spatiotemporally averaged correlation functions after zonal flows have been removed. By integrating these correlation functions with respect to displacements, it is possible to calculate correlation lengths. Figure 5.19 presents $x$ and $y$ correlation lengths for simulations of the cyclone base case at temperature gradient scale length $a/L_T = 5$ at $y$ box sizes $8\pi$, $14\pi$, $28\pi$, $40\pi$, $56\pi$, and $112\pi$ $\rho_i$. Notice that in Figure 5.18, subtracting the zonal mode from the potential results in correlations in $y$ that are negative for a range of displacements. Figure 5.19 shows that directly integrating this function results in a correlation length of zero. For the purposes of developing a correlation heuristic, the sign of the correlation should be irrelevant. As a first pass at a correlation length that accounts for both positive and negative correlations, we choose to integrate the absolute value of the correlation function, defining the correlation length as

$$\ell_{c,x} \equiv \int \mathrm{d}(\Delta x) |\langle \overline{\mathcal{C}_x} \rangle_{xy}|, \tag{5.17}$$

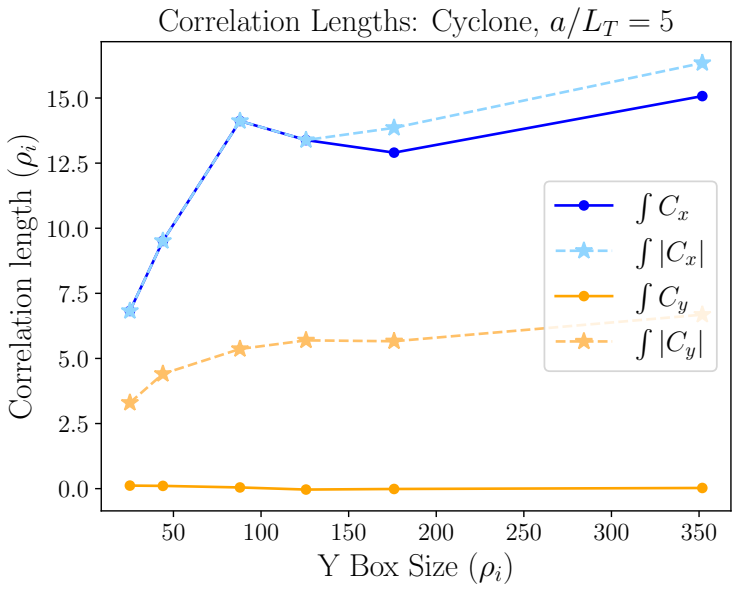



Figure 5.19: Correlation lengths for the cyclone base case as a function of box size in $y$ for normalized temperature gradient $a/L_T = 5$. The blue points correspond to correlations in $x$, while the orange points are for correlations in $y$. Round markers indicate integrals of the correlation function with respect to displacement. Star markers represent integrals of the absolute value of the correlation function, which penalizes the resulting correlation length for potential negative correlations. For the cyclone base case, $x$ correlation lengths are consistently longer than the $y$ correlation lengths after removing the zonal flows.

where we have removed the zonal flow from the potential before calculating the correlation function, and the overline represents a time average. This is the most restrictive correlation length of the four options presented in Figure 5.19, as in all cases we observed, the $x$ correlations are longer than the $y$ correlations. With a correlation length defined, we can now assess the prospects of its use for in indication of improper domain configuration.

We report the results of two physics parameter scans, one over the normalized temperature gradient, and one over the ratio between the electron and ion temperatures, while varying the perpendicular domain size. We use $L_y$ to control $L_x$ though the twist-and-shift boundary condition, and the ratio between the resolution in each perpendicular direction $N_y$ and $N_x$ remains constant. For the temperature gradient scan, we execute simulations at temperature gradients

$a/L_T \in \{2.5, 3.5, 5.0, 10.0\}$ and box sizes $L_y \in \{8\pi, 14\pi, 28\pi, 40\pi, 56\pi, 112\pi\}\rho_i$. We use a higher ratio of $N_y/L_y$ for the cases with the smallest box sizes, as maintaining a consistent ratio led to numerical instabilities. Each simulation uses a temperature ratio of $T_e/T_i = 1$. Figure 5.20 reports the heat fluxes $Q_i$ that result from this scan. Uncertainty in the heat flux is inversely proportional with box size. Considering that the main target applications for large ensembles of electrostatic ITG simulations are optimization, surrogate modeling development, and multiscale transport solver iteration, the additional uncertainty in the heat flux at small box sizes may have a tangible impact on performance, even in cases where the heat flux converges.

Figure 5.21 transforms the data in Figure 5.20 by using (5.17) to normalize the box sizes by the correlation length calculated from each simulation. For the cyclone base case, the correlation length tends to decrease with temperature gradient, which increases the effective box size. An universal threshold for convergence in effective box size would be ideal, as the correlation function is inexpensive to compute. Even after normalizing by correlation length, there is no apparent universal threshold for box size, as increasing the temperature gradient increases the box size required to achieve convergence. For $a/L_T < 5$, the heat flux converges within error bars around an effective box size of ten. However, larger domains are necessary for higher temperature gradients.

Figures 5.22 and 5.23 report heat fluxes for the parameter scan over the temperature ratio $T_e/T_i$ and box sizes $L_y \in \{5\pi, 10\pi, 14\pi, 28\pi, 56\pi, 112\pi\}$. Each simulation has a fixed temperature gradient of $a/L_T = 2.5$. The domain size that is necessary to converge to an accurate heat flux appears to be independent of the temperature ratio, with $L_y = 28\pi$ being sufficient in all cases. An effective box size of $L_y = 5\ell_{c,x}$ produces the correct heat flux.

Finally, we perform a temperature gradient scan for the $\alpha = 0$ field line from the W7-X flux surface in the right side of Figure 5.1. In addition to the benchmark value of $a/L_T = 3.0$, we

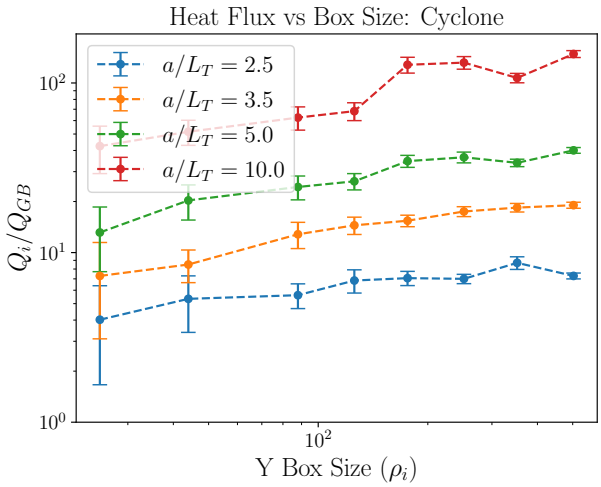



Figure 5.20: Log-log plot of heat fluxes for the cyclone base case for a parameter scan in $y$ box size and temperature gradient. Uncertainty in the heat flux is inversely proportional with box size, despite heat flux increasing with box size.

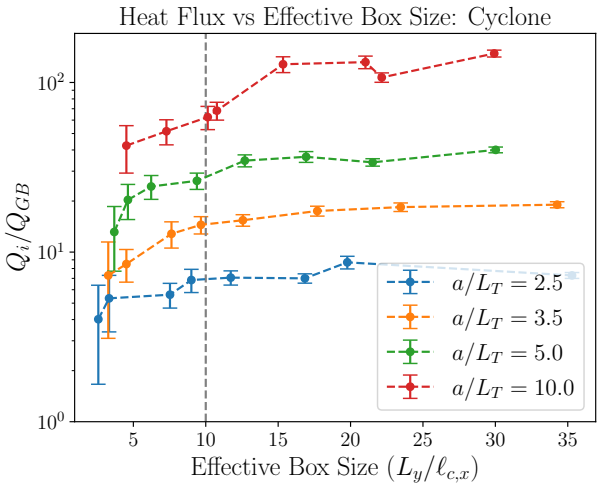



Figure 5.21: Log-linear plot of heat flux as a function of effective box size for the same parameter scan over temperature gradient as in Figure 5.20. For $a/L_T < 5$, the heat flux appears to converge when the effective box size is at least ten times longer than $\ell_{c,x}$, the correlation length in $x$.

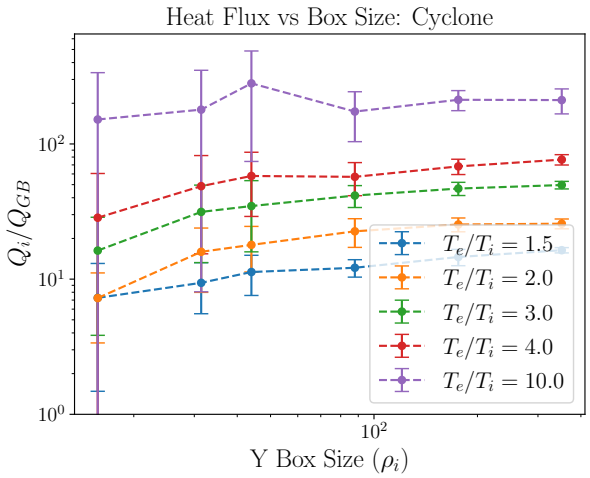



Figure 5.22: Log-log plot of heat flux as a function of box size $L_y$ for the cyclone base case at five values of the ratio between the electron and ion temperatures $T_e/T_i$. Although the heat flux increases with the temperature ratio, the perpendicular domain size requirement appears to be independent of the temperature ratio.

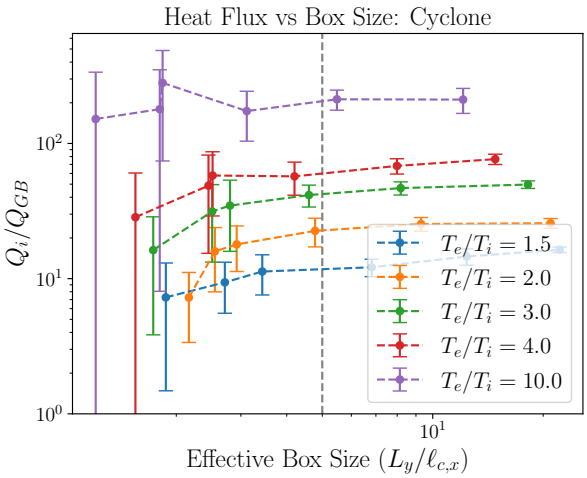



Figure 5.23: Log-log plot of heat flux as a function of box size normalized by the correlation length in $x$. For all values of the temperature ratio, a perpendicular domain that measures roughly five correlation length is sufficient to calculate an accurate heat flux.

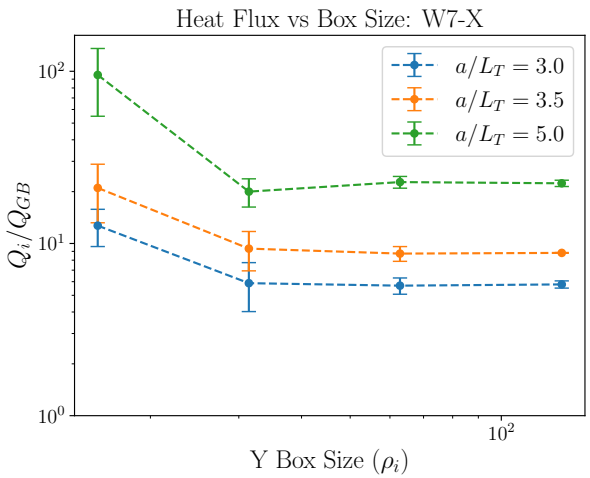



Figure 5.24: Log-log plot of heat flux as a function of box size for a parameter scan in temperature gradient $a/L_T$. When compared to the cyclone base case, the W7-X simulations converge to an accurate heat flux at a smaller box size. Interestingly, a domain that is too small leads to an overestimate of the heat flux.

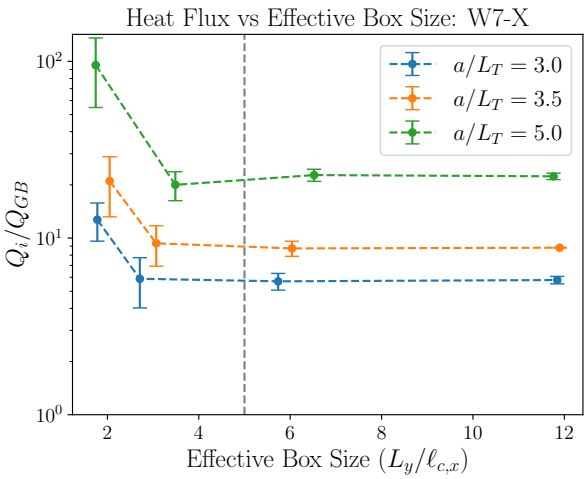



Figure 5.25: Log-linear plot of heat flux compared to effective box size for a scan over $a/L_T$ for the W7-X flux tube. Somewhere between four and five correlation lengths in $x$ seem to yield the converged heat flux.

simulate $a/L_T = 3.5$ and 5.0. Box sizes vary according to $L_y \in \{5\pi, 10\pi, 20\pi, 42\pi\}\rho_i$, where the final value is the benchmark. Figures 5.24 and 5.25 show the relationship between perpendicular box size and heat flux. Increasing the temperature gradient increases the effective box size for $L_y = 10\pi$ and $L_y = 20\pi$, but for $L_y = 42\pi$, the effective box size is constant. Unlike the cyclone base case, a domain that is too small to converge to the correct heat flux leads to an overestimate of the heat flux. A proposed mechanism for that could be that the box may be too small to resolve the zonal mode. While the zonal mode has $k_y = 0$, it has a finite radial extent in $x$ that may be larger than $5\pi\rho_i$. Though this could be investigated further, the main conclusion we draw from this result is that the magnetic geometry impacts whether an improperly sized domain overestimates or underestimates the heat flux.

## 5.3 Resolution

In addition to ensuring that modes have adequate space to be represented, it is also important to provide enough resolution to grant access to fine scales for dissipation. In broad generality, spectra of turbulence can be divided into three groupings of scales, an energy-containing range, an inertial range, and a dissipation range [Davidson, 2015]. As discussed in Chapter 4, recurrence phenomena or numerical reflections can lead to pile-up in energy spectra. To mitigate this effect, simulations typically introduce numerical dissipation, which augments the solution by adding an extra energy sink in a range near the finest scales. This numerical dissipation is not a perfect solution in isolation, as the simulation domain must still resolve the hierarchy of scales present in the inertial range of the turbulence in order to converge to the true solution. A possible problem that could occur would be the direct injection of energy from the energy-containing scales to the dissipation scales, bypassing the inertial range. While turbulence primarily involves local interactions in position and velocity dual spaces, some nonlocal interactions occur through nonlinear terms in the equations. In this chapter, we investigate the potential for turbulence spectra and a nonlinear energy flux diagnostic to indicate whether simulations are adequately resolved.

To support this investigation, we perform a scan over the resolution $N_y$, while maintaining a fixed ratio with $N_x$. We return to the cyclone base case with temperature gradient $a/L_T = 2.5$ and box size $L_y = 56\pi$, which was a well-resolved case at $N_y = 64$. We then vary the resolution such that $N_y \in \{16, 32, 64, 128, 256\}$, which results in equal spacing on a logarithmic scale. For the convergence study, we include additional simulations at $N_y = 24$ and $N_y = 48$ to add more data before the threshold of convergence. Time traces of the heat fluxes of these simulations are presented in Figure 5.26. Qualitatively, the only time trace that appears abnormal is from the

$N_y = 16$ simulation. The large peaks near $t = 375a/v_{ti}$ and $t = 500a/v_{ti}$ may be indications that it has not yet reached statistical steady state. In future work, we could explore this question using the `QUENDS` Python package [Howard et al., 2025], a parallel effort to the work of this thesis that aims to automatically quantify uncertainties in time series data. The other time traces appear reasonable in the absence of context of other simulations at higher resolutions. The $N_y = 32$ simulation underestimates the heat flux by more than fifty percent, but taken in isolation, the time trace shows no indication that the simulation is under-resolved. This is a concrete example of the value of a heuristic for under-resolution. We desire a means of identifying that the simulation is under-resolved without the expense of generating a full ensemble like the one shown in Figure 5.26.

Next, we extend the analysis to include a resolution scan at $a/L_T = 3.5$ and $a/L_T = 5.0$, in addition to the original $a/L_T = 2.5$. Figure 5.27 demonstrates that for the cyclone base case, the lower temperature gradient simulations of $a/L_T = 2.5$ and $a/L_T = 3.5$ require less resolution to produce the correct heat flux than the higher temperature simulations. For this box size, the lower temperature gradient simulations converge by $N_y = 64$, while $a/L_T = 5.0$ requires more resolution. With clear cases of under-resolution easily identifiable in Figure 5.27, we can analyze spectra and energy fluxes in search of clear signals of under-resolution.

### 5.3.1 Analysis of Heat Flux and Free Energy Spectra

To begin our analysis of the spectra, we first define the free energy of the system as

$$W_g = \sum_{\ell,m} \int \mathrm{d}x\mathrm{d}y\mathrm{d}z |G_{\ell,m}|^2, \tag{5.18}$$

and recall the definition for the heat flux (5.16). Heat flux and free energy spectra can be obtained by omitting the integral over $y$ from (5.16) and (5.18), while retaining the native Fourier representation for $h$ and $g$. The justification for this stems from the famous Parseval's theorem for the Fourier transform [Stein and Shakarchi, 2003]:

$$\sum_k \widehat{F}_k \widehat{F}_k^* = \int \mathrm{d}x f^2(x), \tag{5.19}$$

where $\widehat{F}_k$ is the Fourier transform of $f(x)$, $*$ denotes complex conjugation, and the integral is over the one period of $f(x)$. Therefore, omitting the integral over one periodic spatial coordinate is equivalent to retaining its constituent Fourier components. We can thus consider the resulting spectrum as the free energy contained in each scale of the remaining coordinate.

The spectra for the $a/L_T = 2.5$ simulations of the cyclone base case from Figure 5.26 are plotted in Figures 5.28 and 5.29 as functions of binormal wavenumber $k_y$. The $N_y = 16$ and $N_y = 32$ spectra have visible markers of under-resolution, even in the absence of the other simulations for this configuration of physics parameters. Neither of them form an inertial range. For the $N_y = 16$ case, the peak of the spectra is the highest wavenumber. In light of the other simulations, we learn that this is the peak of the energy containing scales, but even without access to those simulations, we can recognize that a simulation with that spectrum is invalid. Similarly, the spectra for $N_y = 32$ are too flat to accurately represent ITG turbulence. A heuristic that incorporates the slope of the energy spectrum on a logarithmic scale may be a worthwhile initial filter. At minimum, the spectrum should not peak at the highest wavenumbers, and it should also not increase in the dissipation range. More research is required to disentangle the numerical hyperdiffusion parameter from the fine scale resolution, but that is beyond the scope of this project.

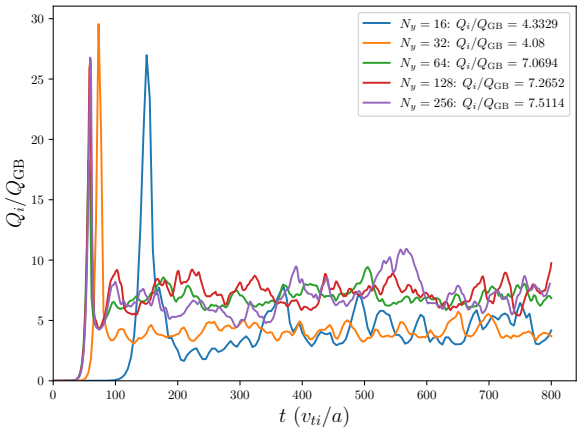



Figure 5.26: Heat flux time traces for five simulations of the cyclone base case at resolutions $Ny \in \{16, 32, 64, 128, 256\}$, all with temperature gradient $a/L_T = 2.5$. The linear instability saturates approximately 100 $t(v_t/a)$ later for $N_y = 16$ than in the high-resolution $N_y = 256$ simulation.

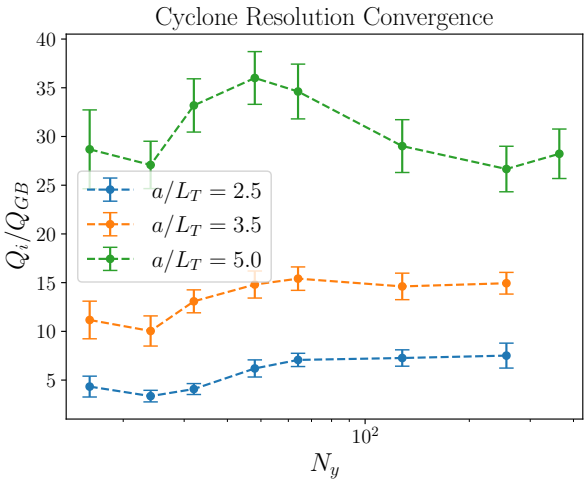



Figure 5.27: Resolution convergence study comparing heat flux to resolution in $y$ for three values of the temperature gradient $a/L_T$. For $a/L_T = 2.5$ and 3.5, the heat flux appears to converge by $N_y = 64$, while more resolution is necessary for $a/L_T = 5$.

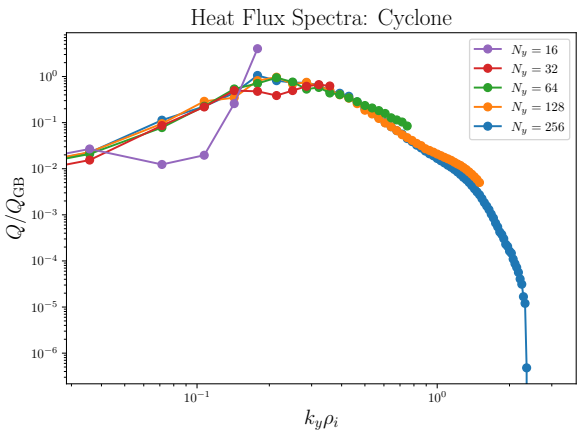



Figure 5.28: Heat flux spectra as a function of $k_y$ for the five simulations of the cyclone base case in Figure 5.26, plotted on a log-log scale. As the low resolution of $N_y = 16$, an inertial range fails to form. The $N_y = 128$ and $N_y = 256$ spectra strongly agree with each other, and the $N_y = 64$ spectrum agrees at low $k_y$.

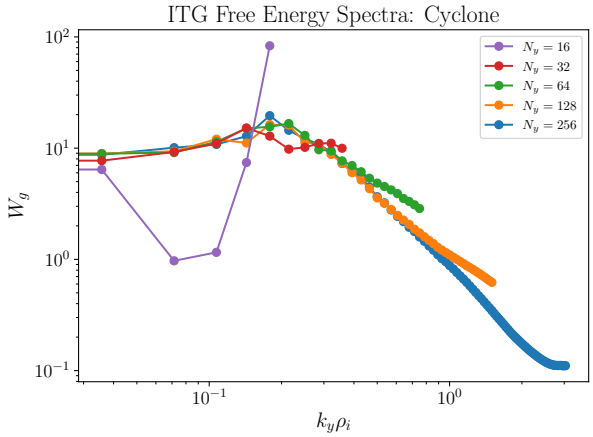



Figure 5.29: Log-log plot of the free energy spectra for the resolution scan shown in Figures 5.26 and 5.28. Unlike the heat flux spectra, the free energy spectrum saturates at high $k_y$ in the $N_y = 256$ case, which may indicate that additional dissipation may be necessary.

### 5.3.2 Nonlinear Energy Transfer Diagnostic

While spectra provide insights into the average energy content of different scales, assessments of the energy fluxes between scales requires a different diagnostic. In this section, we apply the methods of Teaca et al. [2017] to develop an energy flux diagnostic that we can use to study the cyclone base case. In principle, this can be used to assess whether energy is flowing directly from the injection scales to the dissipation scales.

In their paper, Teaca et al. [2017] studied Alfvénic turbulence in the electromagnetic gyrokinetic system in a triply periodic slab to model an astrophysical plasma. They acknowledge that magnetic geometry, in particular shear, complicates the construction of scales. The additional complication introduced by the $z$ dependence of $k_\perp$ led them to focus on slab turbulence. In this

chapter, we study a comparatively simpler physical system in a more complex geometry.

Modifying their definitions for our electrostatic system, we introduce the nonlinear energy transfer function

$$\mathcal{T}(k) \equiv \left\langle \frac{T}{2F_M}(\langle \mathbf{v}_E \rangle_{\mathbf{R}} \cdot \nabla h)h \right\rangle, \tag{5.20}$$

where $T$ is the equilibrium ion temperature, $F_M$ is the background Maxwellian distribution, and $h$ is the non-adiabatic portion of the perturbed distribution function, and they define $\langle ... \rangle$ as a phase space average over the three spatial dimensions and two velocity dimensions. Negative values of the transfer function indicate a net loss of energy at a given scale, and positive values represent a gain of energy.

Next, we formalize the concept of perpendicular scales and their interactions. Because the nonlinear term in the gyrokinetic system is quadratic, nonlinear interactions occur between triads of modes [Teaca et al., 2017],

$$\mathbf{k} + \mathbf{p} + \mathbf{q} = 0, \tag{5.21}$$

where each of these wave vectors represents a perpendicular scale. In this section, to spare notation and to align with the original work, we write $k_\perp$ as $k$. The characteristic length scale of wave-vector $k$ is the inverse of its norm, $1/k$. Due to magnetic geometry,

$$k(z) = \sqrt{|\nabla x|^2 k_x^2 + 2\nabla x \cdot \nabla y k_x k_y + |\nabla y|^2 k_y^2}, \tag{5.22}$$

where $\nabla x$ and $\nabla y$ are functions of $z$. Figure 5.30 illustrates how magnetic shear couples different pairs of $k_x$ and $k_y$ wavenumbers together at different points along the field line. At the outboard midplane with $\theta = 0$, the wavebands form concentric half-circular shells. At $\theta = \pi/2$, the wave-

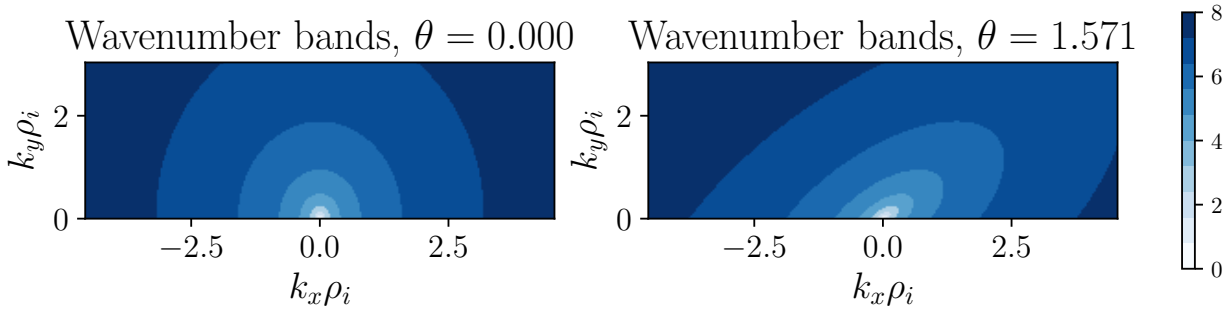



Figure 5.30: Perpendicular waveband shells for the cyclone base case evaluated at $\theta = 0$ (left) and $\theta = \pi/2$ (right). Magnetic shear couples different $(k_x, k_y)$ pairs together at different points along the field line.

bands tilt under the influence of magnetic shear.

Using the triangle inequality, one can show that if two scales $k$ and $p$ of comparable magnitude interact with each other, the third scale $q$ of the triad must be between zero and twice that scale [Teaca et al., 2017]:

$$k \approx p \implies 0 \leq q \leq 2k. \tag{5.23}$$

This motivates their use of a decomposition into a series of wavebands, $s_n = [k_{n-1}, k_n]$ separated by a progression of the form

$$k_n = k_1 \lambda^{n-1}. \tag{5.24}$$

We choose $\lambda = 2$ for a dyadic progression. We lose some fine-scale details in the transfer function this way, but the computational cost of the calculation decreases.

We are interested in calculating the energy flux from waveband $P$ to waveband $K$, which is defined as

$$\mathcal{P}(K|P) \equiv \left\langle \frac{T}{2F_M} (\langle \mathbf{v}_E \rangle_{\mathbf{R}} \cdot \nabla h^{[P]}) h^{[K]} \right\rangle. \tag{5.25}$$

This requires the following steps:

1. Use the quasineutrality condition to calculate the electrostatic potential $\Phi$.

2. Calculate the $\mathbf{E} \times \mathbf{B}$ velocity.

3. Filter $h$ with waveband masks to calculate $h^{[P]}$ and $h^{[K]}$.

4. Average over phase space (5.25).

The waveband masks can be precomputed during initialization. The current implementation of this algorithm is a post-processing script in Python. Because this diagnostic requires the full distribution function, we added a feature to `GX` to store the distribution function to disk at a specified cadence.

The results of applying our implementation of the energy transfer diagnostic to our five cyclone base case simulations from Figure 5.26 are displayed in Figures 5.31 – 5.35. Every square in each image represents a pair of wavebands $P$ and $K$. None of the figures show evidence of significant energy flows from large scales to small scales. Figure 5.31 highlights the peak of the transfer function at the two highest wavebands, which should not occur in well-resolved simulations. The highest resolution simulations in Figures 5.34 and 5.35 show the development of an inertial range, where the majority of the waveband pairs above the diagonal share the same negative sign, indicating that energy is flowing from large scales to small scales. The reversal in sign at the final pairs for $N_y \in \{16, 32, 64\}$ may be evidence of reflections at the boundary that cause the exponent of their corresponding free energy spectra in Figure 5.29 to be higher than in the higher resolution cases. Additional study is necessary to form firm conclusions about these data. As a first pass, a heuristic that analyzes the spectra may be a valuable feature to prioritize, and the more expensive and more detailed nonlinear energy flux diagnostic could be used at a later date.

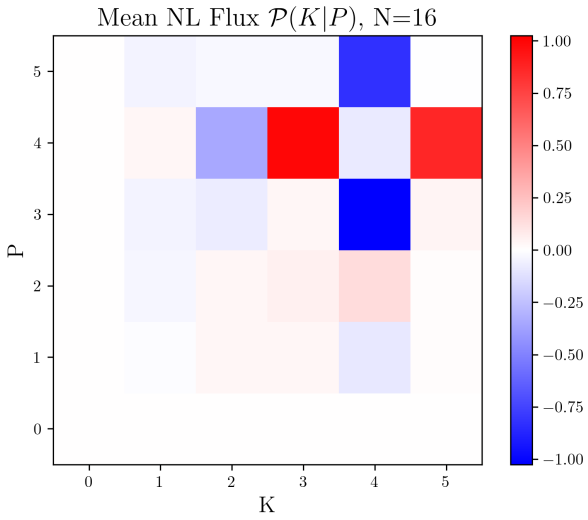



Figure 5.31: Nonlinear energy transfer function $\mathcal{P}(K|P)$ for the $N_y = 16$ simulation of the cyclone base case with $a/L_T = 2.5$ and wavebands spaced by the dyadic progression $k_{n+1} = 2k_n$.

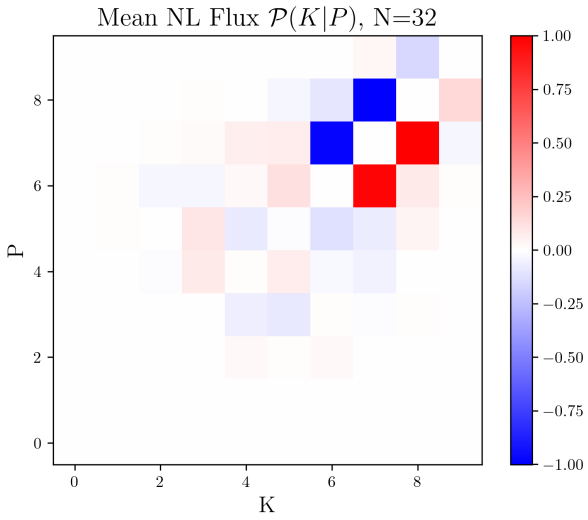



Figure 5.32: Nonlinear energy transfer function $\mathcal{P}(K|P)$ for the $N_y = 32$ simulation of the cyclone base case with $a/L_T = 2.5$.

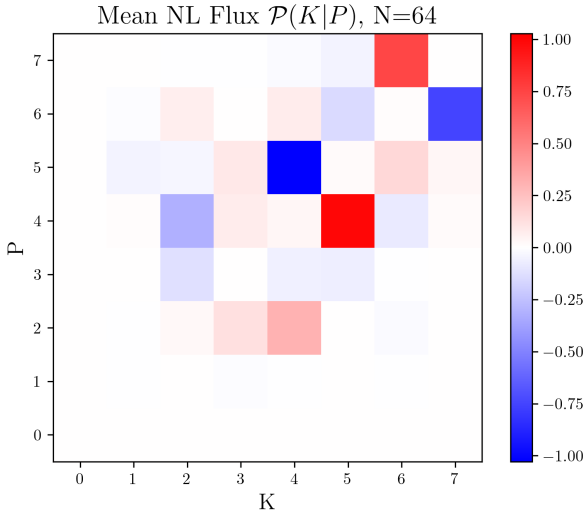



Figure 5.33: Nonlinear energy transfer function $\mathcal{P}(K|P)$ for the $N_y = 64$ simulation of the cyclone base case with $a/L_T = 2.5$.

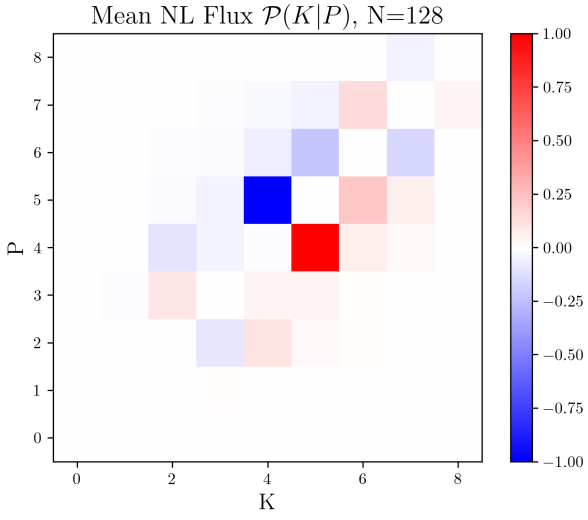



Figure 5.34: Nonlinear energy transfer function $\mathcal{P}(K|P)$ for the $N_y = 128$ simulation of the cyclone base case with $a/L_T = 2.5$.

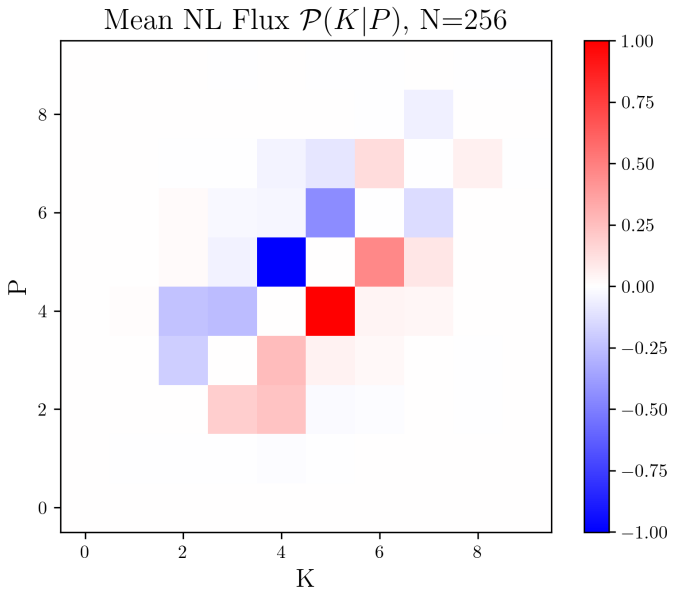



Figure 5.35: Nonlinear energy transfer function $\mathcal{P}(K|P)$ for the $N_y = 256$ simulation of the cyclone base case with $a/L_T = 2.5$.

The nonlinear energy transfer spectra that Teaca et al. [2017] developed for electromagnetic Alfvén turbulence formed a prominent direct energy cascade across almost all of the domain. Moreover, the bottom left region of the plots were well-populated, which corresponded to significant local transfers of energy between the largest resolved scales in the system. In contrast, for all resolutions of these electrostatic simulations of the cyclone base case with adiabatic electrons plotted in Figures 5.31−5.35, the largest scales do not show any contribution to the nonlinear energy transfer function. This discrepancy could be a result of differences in the wavenumber band construction. Teaca et al. [2017] used a fine structure for their wavenumber bands, with $\lambda = 2^{1/3}$. We used less expensive, coarser set of wavenumber bands, with $\lambda = 2$. This choice was motivated by a desire to ensure that all wavenumber bands contained at least one $(k_x, k_y)$ pair. At low resolutions $(N_y = 16)$, the fine $\lambda = 2^{1/3}$ spacing between bands used by Teaca et al. [2017] results in gaps in the transfer spectrum.

Differences between electromagnetic solar wind turbulence and electrostatic, zonal-flow

driven tokamak turbulence could also contribute to the differences between the transfer functions we calculate and those in the original study. As previously discussed, the cyclone base case has magnetic shear that tilts the wavenumber bands at points where $\theta > 0$. Simulations with kinetic electrons provide another channel for possible energy dissipation, as the system gains access to electron scales that are not present in ITG simulations with adiabatic electrons. These physics differences may contribute to the discrepancies between our work and the results of Teaca et al. [2017].

A similar analysis of the advection of free energy via the quadratic nonlinear term has been studied for the cyclone base case [Bañón Navarro et al., 2011]. In their analysis, Bañón Navarro et al. [2011] demonstrate that this system has a preeminent direct cascade of free energy due to a dominant entropy cascade. However, the electrostatic potential portion of the free energy exhibits nonlocal energy transfers. Importantly, they show that dissipation terms are significant at all $k_\perp$, which diverges from the classic picture of dissipation as only a small-scale phenomenon. One possible extension of the work in this chapter would be to verify the results against the exact system studied by Bañón Navarro et al. [2011]. This would require the implementation of a separate entropy and electrostatic energy transfer diagnostic. A dissipation diagnostic that calculates the total energy dissipated at each perpendicular wavenumber scale would also be necessary. We would also use their novel variable wavenumber band scaling, which uses coarse bands at small wavenumbers and finer bands at large wavenumbers.

# Chapter 6: Summary and Discussion

As the fusion industry approaches the inflection point where it prepares to conduct reactor design studies, gyrokinetic simulations are increasingly valuable tools for predictive modeling of turbulent heat transport. An important objective is to accelerate these simulations to achieve accurate results in less time, without compromising the physics fidelity of the models. The projects presented in this thesis represent initial steps toward that goal.

In Chapter 2, we presented a first-principles derivation of the electrostatic gyrokinetic system and the 1-D Vlasov-Poisson system. Though the Vlasov-Poisson system is a simplified model, its quadratic nonlinear term permits it to exhibit nonlinear dynamics that challenges closure models. This feature, along with linear phase mixing, made it a suitable test problem for the electrostatic gyrokinetic system.

In Chapter 3, we introduced our implementation of an Eulerian pseudo-spectral method to solve the electrostatic gyrokinetic and Vlasov-Poisson systems. We discussed the efficiency advantages of pseudo-spectral methods relative to finite difference and other finite element methods. We also derived the projection of the electrostatic gyrokinetic system onto the Fourier-Laguerre-Hermite basis used in the `GX` code. We concluded by projecting the Vlasov-Poisson system on a Fourier-Hermite basis. After applying the appropriate normalizations, both systems share the same time derivative and parallel streaming terms.

Chapter 4 introduced a machine learning closure for parallel velocity in the Vlasov-Poisson system, implemented using a Fourier-Hermite basis. We developed a Python code to solve these equations in three scenarios: a driven-dissipative linear system, a low-amplitude initial-value problem, and a high-amplitude, nonlinear initial-value problem. The ML closure model successfully

reduced velocity-space resolution requirements in all cases, while maintaining accurate and stable evolution of the system.

The eventual goal of this work is to apply the ML closure method to the gyrokinetic system. A step toward that goal would be to implement a transfer learning scheme to demonstrate that the ML closure generalizes to problems outside its training set. While an advantage of reservoir computing is that training is inexpensive relative to other recurrent neural network frameworks, minimizing computational costs at fixed accuracy remains of paramount importance. A possible path toward transfer learning for this test problem could be to train the network on an ensemble of simulations with a variety of initial conditions.

Once transfer learning has been demonstrated, the next step would be to develop an efficient closure for configuration space. Typical simulations of gyrokinetic turbulence in fusion plasmas require at least an order of magnitude more resolution in configuration space than in velocity space. Extending the closure to Fourier space is a more challenging endeavor than developing the velocity-space closure, as the nonlinear term couples all wavenumbers to each other. A naive implementation could be to approximate the small spatial scales with an ML prediction of a single wavenumber. A more sophisticated approach could be to use a more comprehensive training process that directly incorporates the free energy spectrum as a guide for training.

After implementing these improvements to the method, the final step would be to apply the method to the gyrokinetic system. The methodology for using transfer learning on the Fourier-Hermite representation of the Vlasov-Poisson system would follow a similar form for the higher-dimensional gyrokinetic system. Achieving a robust, generalized ML closure for the gyrokinetic system may be more difficult due to the greater number of physics parameters in gyrokinetics.

In Chapter 5, we implemented correlation functions, heat flux and energy spectra, and non-

linear energy transfer diagnostics to study electrostatic gyrokinetic turbulence. We used GX to perform parameter scans over the temperature gradient and temperature ratio parameters for a tokamak and a stellarator case. In tandem, we varied the domain extent and resolution in the perpendicular plane in search of a simple heuristic that could be used to determine whether simulations require a larger domain or finer grid to produce an accurate heat flux. Though we did not find a universal threshold for either correlation lengths or resolution, the data exhibited thresholds for each case. The more sophisticated nonlinear energy transfer diagnostic did not show clear trends at low resolution, but the high-resolution cases showed intuitive indications of an inertial range. In future work, these diagnostics can be directly integrated into GX during runtime to evaluate ongoing simulations. For some cases, heat flux and free energy spectra showed clear signs of poor resolution, and which coincided with inaccurate heat fluxes. A preliminary version of a resolution evaluation tool could be developed to monitor these spectra for signs of under-resolution.

There are an abundance of avenues for continued effort toward achieving automatic convergence testing. Though the ITG simulations we performed for W7-X stellarator showed minimal correlation lengths along the magnetic field line, we expect that there is a relevant minimum length requirement for the flux tubes. To test our field period symmetry explanation for the short parallel correlation lengths in the W7-X case, we could repeat the analysis for a magnetic equilibrium for a different stellarator with a lower number of field periods. We predict that this would result in longer parallel correlation lengths than the W-7X case.

Additional simulations are needed to develop a comprehensive threshold for insufficient perpendicular domain size. The W7-X benchmark required a larger perpendicular box at high temperature gradients than the cyclone base case required. Calculating perpendicular correlation lengths for ITG simulations of a broader range of magnetic geometries may yield insights into the prospects

of using correlation lengths as a reduced metric for proper domain specification. Similarly, other physics parameters should be scanned. Staying within the locus of ITG simulations with adiabatic electrons, we could vary the density gradient, rotational transform, magnetic shear, and the ratio of $L_y/L_x$ to investigate their impacts on the link between correlation lengths and accurate heat fluxes. Using correlation lengths as criteria for proper perpendicular domain configuration would be ideal, as comparing scalars is more straightforward than comparing arrays like spectra.

A potentially interesting avenue to explore for the automatic resolution convergence portion of this work would be to directly use the turbulent heat flux and free energy spectra to diagnose convergence. In principle, one could leverage contemporary advances in computer vision and artificial intelligence and cast this task as an image classification problem. Given several spectra, a successful algorithm would classify a simulation as well-resolved or poorly resolved. Training could involve the creation of a dataset of ITG simulations that span a representative range of parameters and geometries, each performed at multiple resolutions. While building the dataset of simulations would be expensive, the process could be the result of the passive, gradual accumulation of simulations that originally were used for other studies. If this approach to solving the problem seems too cumbersome or costly, another method would be to establish a simpler heuristic. We could implement automatic checks for pileup in either the heat flux or free energy spectra. Hermite and Laguerre spectra could also be used to check velocity space convergence, similar to our approach in Chapter 4 for the ML closure study.

Finally, the nonlinear energy transfer diagnostic could yield supplemental information that spectra cannot capture alone. The immediate objective would be to benchmark the method against the published work of Bañón Navarro et al. [2011] for the cyclone base case. Then, we could implement a coarser diagnostic that could identify whether energy is passed from injection scales

directly to dissipation scales. This would require us to create a diagnostic that measures the total dissipation from a band of wavenumbers. The free energy spectra are already a diagnostic that measures the energy content of each scale, so they could be used to define the energy containing scales. Once this reduced form of the nonlinear energy transfer diagnostic is implemented, the next step would be to perform another parameter scan to assess its predictive power.

# Appendix A: Appendix to ML Closure Study

## A.1 Hermite Resolution Convergence

In Section 4.3.3, we demonstrated that reservoir computing can be used to construct an ML closure that reduces number of moments required to capture the low-order spectrum of the strongly nonlinear regime of the Vlasov-Poisson system by a factor of sixteen. In Figure A.1, we present the convergence study that we used to determine the high-resolution ($M = 65$) baseline for that case. The low-order moments represent important physical quantities, including density, momentum, and energy. Therefore, accurately resolving those moments is a higher priority than resolving the fine-scale structure in velocity-space that the high-order moments capture. In Figure A.1, the lowest Hermite resolution that captures the $m = 0$ and $m = 1$ density and momentum moments within a factor of 2 is $M = 65$, so that resolution was selected as the baseline case. Figure A.2 shows a root-mean-square error (RMSE) metric for the low-order Hermite spectra at different Hermite resolutions, where we take the $M = 513$ spectrum as ground truth. Inaccuracy in the density and momentum moments leads to increased RMSE for the $M = 17$ and $M = 33$ cases, and subsequent errors converge exponentially, beginning with $M = 65$.

## A.2 Time Evolution of Low-Order Moments

Sections 4.3.2 and 4.3.3 demonstrate that the ML closure can capture spectra of the Vlasov-Poisson system and that the time evolution of the density moment is accurate. Figures A.3 – A.6

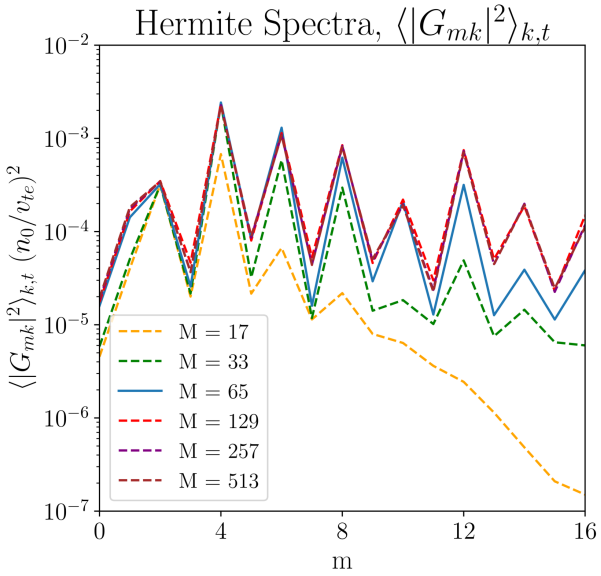



Figure A.1: Convergence study of the Hermite spectra for the strongly nonlinear initial-value problem from Section 4.3.3. We solve (3.94) - (3.97) without the ML closure at increasing levels of resolution in velocity space. We selected $M = 65$ moments for the high-resolution baseline case in Section 4.3.3, as that case is the first to show convergence in the density and momentum moments.

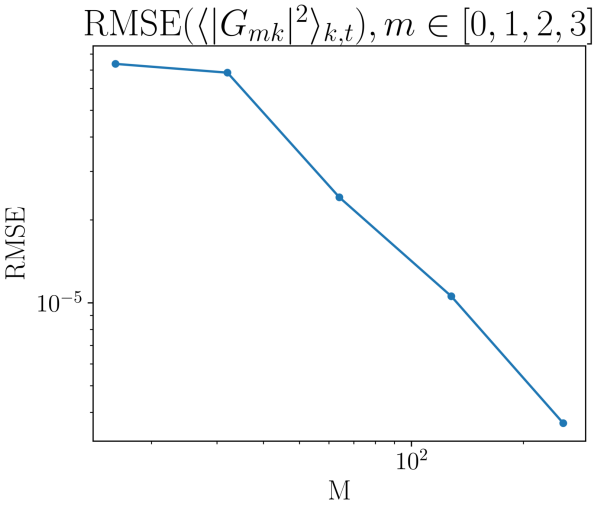



Figure A.2: Root-mean-square error in the low-order ($m \in [0, 1, 2, 3]$) Hermite spectra at the resolutions plotted in Figure A.1, as compared to the $M = 513$ case. The $M = 17$ and $M = 33$ cases have similar errors, and exponential convergence occurs afterward.

| $\epsilon$ | $m$ | **DNS damping rate** $(\omega_{pe})$ | **ML closure damping rate** $(\omega_{pe})$ |
|---|---|---|---|
| 0.001 | 1 | $-6.581 \times 10^{-2}$ | $-6.911 \times 10^{-2}$ |
| 0.001 | 2 | $-6.581 \times 10^{-2}$ | $-6.911 \times 10^{-2}$ |
| 0.18 | 1 | $-5.476 \times 10^{-3}$ | $-7.572 \times 10^{-3}$ |
| 0.18 | 2 | $-8.732 \times 10^{-3}$ | $-7.518 \times 10^{-3}$ |

Table A.1: Damping rates calculated from the initial-value simulations in Figures A.3 – A.6 . Here, $\epsilon$ is the amplitude of the initial cosine density perturbation, and $m$ is the Hermite moment number. Damping rates were calculated by solving a linear regression problem for the natural logarithm of the local maxima in each time series.

depict time traces of the $m = 1$ and $m = 2$ moments for the cases in those sections. The ML closure resolves the frequencies of the oscillations well. It also successfully captures the amplitude peaks soon after training ends, but it fails to match many of the minima. At long times after training, the amplitudes resulting from the ML closure begin to deviate slightly from the DNS amplitudes. In the weakly nonlinear regime presented in Figures A.3 and A.4 , the ML closure leads to a slightly lower amplitude at the end of the simulation for both the $m = 1$ and $m = 2$ moments. In the strongly nonlinear regime in Figure A.5 , the ML closure trends toward a lower amplitude for $m = 1$ than in the DNS baseline. However, for $m = 2$ in that regime, Figure A.6 shows that ML closure trends toward a slightly higher amplitude than the DNS. To quantitatively analyze this behavior, we report damping rates for the DNS and ML closure in table A.1 . Damping rates were calculated by first extracting the local maxima from each time trace and then solving a linear regression problem for the natural logarithm of those maxima. An interesting observation is that the $m = 1$ and $m = 2$ results are identical for the low-amplitude case. In contrast, in the high-amplitude case, the DNS has a faster damping rate for $m = 2$ than for $m = 1$. In the strongly nonlinear regime, the ML closure yields similar damping rates for $m = 1$ and $m = 2$, but they diverge from the DNS damping rates.

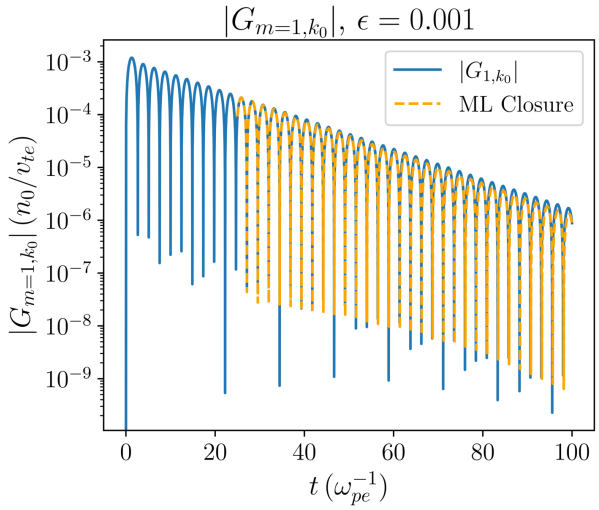



Figure A.3: Time trace of the $m = 1$ moment for the low-amplitude initial-value case in Section 4.3.2 .

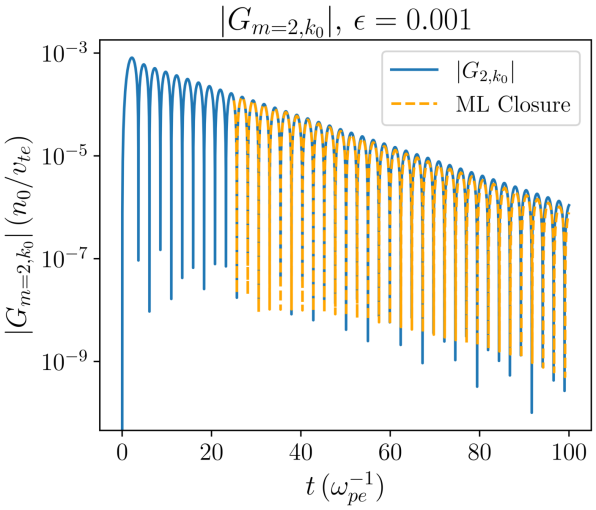



Figure A.4: Time trace of the $m = 2$ moment for the low-amplitude initial-value case in Section 4.3.2 .

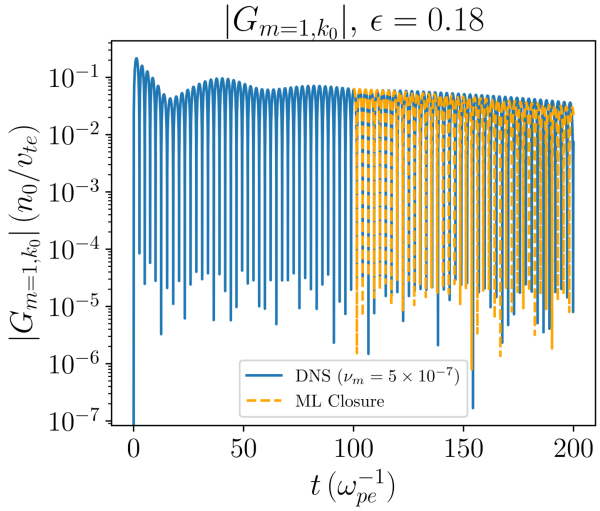



Figure A.5: Time trace of the $m = 1$ moment for the high-amplitude initial-value case in Section 4.3.3 .

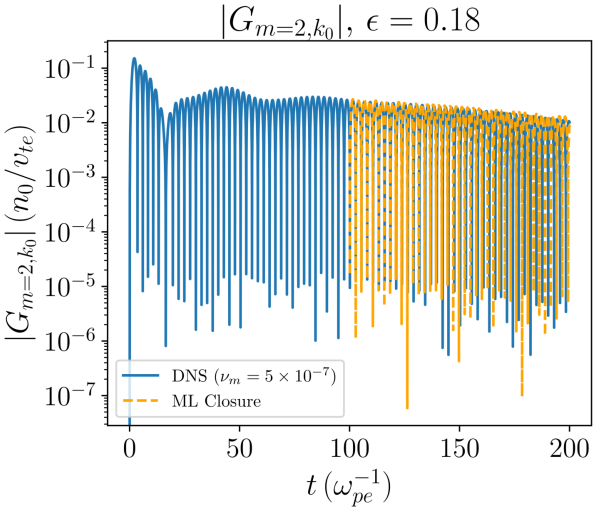



Figure A.6: Time trace of the $m = 2$ moment for the high-amplitude initial-value case in Section 4.3.3 .

# Bibliography


118th United States Congress. Fire grants and safety act, 2024. Sec. 205: Fusion Energy Regulation.

I. G. Abel and S. C. Cowley. Multiscale gyrokinetics for rotating tokamak plasmas: II. Reduced models for electron dynamics. *New. J. Phys.*, 15:023041, 2013.

Ian G. Abel, Gabriel G. Plunk, E. Wang, Michael Barnes, Steven C. Cowley, William Dorland, and Alexander A. Schekochihin. Multiscale gyrokinetics for rotating tokamak plasmas: fluctuations, transport and energy flows. *Rep. Prog. Phys.*, 76:116201, 2013.

Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, pages 773–782. U.S. Dept. of Commerce, National Bureau of Standards; Dover Publications, tenth edition, 1972.

Toby Adkins and A. A. Schekochihin. A solvable model of Vlasov-kinetic plasma turbulence in Fourier-Hermite phase space. *J. Plasma Phys.*, 84(1):905840107, 2018.

Ali Q. Al-Shetwi, Izham Zainal Abidin, Khaled A. Mahafzah, and M.A. Hannan. Feasibility of future transition to 100% renewable energy: Recent progress, policies, challenges, and perspectives. *J. Cleaner Production*, 478:143942, 2024.

Othman Abdullah Almatroud, Karthikeyan Rajagopal, Viet-Thanh Pham, and Giuseppe Grassi. A novel chaotic system with only quadratic nonlinearities: Analysis of dynamical properties and stability. *Mathematics*, 12:612, 2024.

Thomas M. Antonsen, Jr. and Barton Lane. Kinetic equations for low frequency instabilities in inhomogeneous plasmas. *Phys. Fluids*, 23:1205–1214, 1980.

Troy Arcomano, Istvan Szunyogh, Jaideep Pathak, Alexander Wikner, Brian R. Hunt, and Edward Ott. A machine learning-based global atmospheric forecast model. *Geophys. Research Lett.*, 47: e2020GL087776, 2020.

George B. Arfken, Hans J. Weber, and Frank E. Harris. *Mathematical Methods for Physicists*, chapter 18. Elsevier, seventh edition, 2013.

Thomas P. Armstrong. Numerical studies of the nonlinear Vlasov equation. *Phys. Fluids*, 10(6): 1269–1280, 1967.

A. Y. Aydemir. A unified monte carlo interpretation of particle simulations and applications to non-neutral plasmas. *Phys. Plasmas*, 1:822–831, 1994.

A. Bañón Navarro, P. Morel, M. Albrecht-Marc, D. Carati, F. Merz, T. Görler, and F. Jenko. Free energy cascade in gyrokinetic turbulence. *Phys. Rev. Lett.*, 106:055001, 2011.

Alejandro Bañón Navarro, Bogdan Teaca, Frank Jenko, Gregory W. Hammett, T. Happel, and ASDEX Upgrade Team. Applications of large eddy simulation methods to gyrokinetic turbulence. *Phys. Plasmas*, 21:032304, 2014.

Scott D. Baalrud and Jérôme Daligault. Effective potential kinetic theory for strongly coupled plasmas. *AIP Conference Proceedings*, 1786:130001, 2016.

Nathaniel Barbour, William Dorland, Ian George Abel, and Matt Landreman. Machine–learning closure for Vlasov–Poisson dynamics in Fourier–Hermite space. *J. Plasma Phys.*, 91:E140, 2025.

Michael Barnes, Felix I. Parra, and Alexander A. Schekochihin. Critically balanced ion temperature gradient turbulence in fusion plasmas. *Phys. Rev. Lett.*, 107:115003, 2011.

Michael Barnes, Felix I. Parra, and Matt Landreman. `stella`: An operator-split, implicit-explicit $\delta f$-gyrokinetic code for general magnetic field configurations". *J. Comput. Phys.*, 391:365–380, 2019.

Michael Barnes, David Dickinson, William Dorland, Peter Alec Hill, Joseph Thomas Parker, Colin Malcolm Roach, Maurizio Giacomin, Noah Mandell, Ryusuke Numata, Tobias Schuett, Stephen Biggs-Fox, Nicolas Christen, Jason Parisi, George Wilkie, Lucian Anton, Justin Ball, Jessica Baumgaertel, Greg Colyer, Michael Hardman, Joachim Hein, Edmund Highcock, Gregory Howes, Adrian Jackson, Michael T. Kotschenreuther, Jungpyo Lee, Huw Leggate, Adwiteey Mauriya, Bhavin Patel, Tomo Tatsuno, and Ferdinand Van Wyk. `GS2` v8.2.1, 2024. URL `https://doi.org/10.5281/zenodo.14022029`.

M. A. Beer, S. C. Cowley, and G. W. Hammett. Field-aligned coordinates for nonlinear simulations of tokamak turbulence. *Phys. Plasmas*, 2:2687–2700, 1995.

Hans A. Bethe. Energy production in stars. *Phys. Rev.*, 55, 1939.

C. K. Birdsall and A. B. Langdon. *Plasma Physics via Computer Simulation*. CRC Press, 1991.

Ludwig Boltzmann. Weitere studien über das wörmegleichgewicht unter gasmolekülen. *Sitzungsberichte Akad. Wiss.*, 66:275–370, 1872. English translation in *The Kinetic Theory of Gases*, Imperial College Press (2003).

J. P. Boyd. *Chebychev and Fourier Spectral Methods*, pages 1–17. Dover Publications, second edition, 2000.

M. Brunetti, F. Califano, and F. Pegoraro. Asymptotic evolution of nonlinear Landau damping. *Phys. Rev. E*, 62(3):4109, 2000.

J. D. Callen, C. C. Hegna, and A. J. Cole. Transport in tokamak plasmas. *Phys. Plasmas*, 17: 056113, 2010.

J. Candy and R. E. Waltz. An Eulerian gyrokinetic-Maxwell solver. *J. Comp. Phys.*, 186:545, 2003.

J. Candy, E.A. Belli, and R.V. Bravenec. A high-accuracy Eulerian gyrokinetic solver for collisional plasmas. *J. Comput. Phys.*, 324:73–93, 2016.

M. P. Carpentier and A. F. Dos Santos. Solution of equations involving analytic functions. *J. Comput. Phys.*, 45(2):210, 1982.

J. Castagna, F. Schiavello, L. Zanisi, and J. Williams. StyleGAN as an AI deconvolution operator for large eddy simulations of turbulent plasma equations in BOUT++. *Phys. Plasmas*, 31: 033902, 2024.

Peter J. Catto. Linearized gyro-kinetics. *Plasma Phys.*, 20:719, 1978.

Ashesh Chattopadhyay, Pedram Hassanzadeh, and Devika Subramanian. Data-driven predictions of a multiscale Lorenz 96 chaotic system using machine-learning methods: reservoir computing, artificial neural network, and long short-term memory network. *Nonlin. Processes Geophys.*, 27: 373–389, 2020.

Wenjie Cheng, Haiyang Fu, Liang Wang, Chuanfei Dong, Yaqiu Jin, Mingle Jiang, Jiayu Ma, Yilan Qin, and Kexin Liu. Data-driven, multi-moment fluid modeling of Landau damping. *Computer Phys. Communications*, 282:108538, 2023.

Benoit Clavier, David Zarzoso, Diego del-Castillo-Negrete, and E. Frénod. Generative-machine-learning surrogate of plasma turbulence. *Phys. Rev. E*, 111:L013202, 2025.

James W. Cooley and John W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comp.*, 19:297–301, 1965.

P. A. Davidson. *Turbulence: an introduction for scientists and engineers*, chapter 3. Oxford University Press, second edition, 2015.

James W. Deardorff. A numerical study of three-dimensional turbulent channel flow at large Reynolds numbers. *J. Fluid Mech.*, 41:453–480, 1970.

Jonas Degrave, Federico Felici, Jonas Buchli, Michael Neunert, Brendan Tracey, Francesco Carpanese, Timo Ewalds, Roland Hafner, Abbas Abdolmaleki, Diego de las Casas, Craig Donner, Leslie Fritz, Cristian Galperti, Andrea Huber, James Keeling, Maria Tsimpoukelli, Jackie Kay, Antoine Merle, Jean-Marc Moret, Seb Noury, Federico Pesamosca, David Pfau, Olivier Sauter, Cristian Sommariva, Stefano Coda, Basil Duval, Ambrogio Fasoli, Pushmeet Kohli, Koray Kavukcuoglu, Demis Hassabis, and Martin Riedmiller. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature*, 602:414–419, 2022.

William Denis D'haeseleer, William Nicholas Guy Hitchon, James D. Callen, and J. Leon Shohet. *Flux Coordinates and Magnetic Field Structure*. Springer, 1991.

A. M. Dimits, G. Bateman, M. A. Beer, B. I. Cohen, W. Dorland, G. W. Hammett, C. Kim, J. E. Kinsey, M. Kotschenreuther, A. H. Kritz, L. L. Lao, J. Mandrekas, W. M. Nevins, S. E. Parker, A. J. Redd, D. E. Shumaker, R. Sydora, and J. Weiland. Comparisons and physics basis of tokamak transport models and turbulence simulations. *Phys. Plasmas*, 7:969–983, 2000.

G. Dimonte and J. H. Malmberg. Destruction of trapping oscillations. *Phys. Fluids*, 21:1188–1206, 1978.

John Donaghy and Kai Germaschewski. In search of a data-driven symbolic multi-fluid ten-moment model closure. *J. Plasma Phys*, 89:895890105, 2023.

W. Dorland and G. W. Hammett. Gyrofluid turbulence models with kinetic effects. *Phys. Fluids B*, 5:812, 1993.

William Dorland, Frank Jenko, Mike Kotschenreuther, and Barett N. Rogers. Electron temperature gradient turbulence. *Phys. Rev. Lett.*, 85:5579, 2000.

J. P. Dougherty. Model Fokker-Planck equation for a plasma and its solution. *Phys. Fluids*, 7: 1788–1799, 1964.

Dale R. Durran. *Numerical Methods for Fluid Dynamics*, chapter 2. Springer, second edition, 2010.

Mostafa EsmaeiliShayan. Solar energy storage systems: A comprehensive study for techno-economic aspects and sustainable grid integration. *Journal of Cleaner Production*, 529:146811, 2025.

Baptiste J. Frei, Antoine C. D. Hoffmann, Paolo Ricci, Stephan Brunner, and Zeno Tecchioll. Moment-based approach to the flux-tube linear gyrokinetic model. *J. Plasma Phys.*, 89: 905890414, 2023.

Baptiste J. Frei, J. Mencke, and Paolo Ricci. Full-F turbulent simulation in a linear plasma device using a gyro-moment approach. *Phys. Plasmas*, 31:012301, 2024.

Jeffrey P. Freidberg. *Plasma Physics and Fusion Energy*, chapter 2. Cambridge University Press, 2007.

E. A. Frieman and Liu Chen. Nonlinear gyrokinetic equations for low-frequency electromagnetic waves in general plasma equilibria. *Phys. Fluids*, 25:502–508, 1982.

Yichen Fu, David Eldon, Keith Erickson, Kornee Kleiwegt, Leonard Lupin-Jimenez, Mark D Boyer, Nick Eidietis, Nathaniel Barbour, Olivier Izacard, and Egemen Kolemen. Machine learning control for disruption and tearing mode avoidance. *Phys. Plasmas*, 27(2):022501, 2020.

X. Garbet, Y. Idomura, L. Villard, and T.H. Watanabe. Gyrokinetic simulations of turbulent transport. *Nucl. Fusion*, 50:043002, 2010.

J Geiger, C D Beidler, Y Feng, H Maaßberg, N B Marushchenko, and Y Turkin. Physics in the magnetic configuration space of W7-X. *Plasma Phys. and Controlled Fusion*, 57:014004, 2015.

K. Germaschewski, B. Allen, T. Dannert, M. Hrywniak, J. Donaghy, G. Merlo, S. Ethier, E. D'Azevedo, F. Jenko, and A. Bhattacharjee. Toward exascale whole-device modeling of fusion devices: Porting the `GENE` gyrokinetic microturbulence code to GPU. *Phys. Plasmas*, 28:062501, 2021.

Livio Gibelli and Bernie D. Shizgal. Spectral convergence of the Hermite basis function solution of the Vlasov equation: The free-streaming term. *J. Comput. Phys.*, 219:477–488, 2006.

P. Goldreich and S. Sridhar. Toward a theory of interstellar turbulence. II. Strong Alfvénic turbulence. *Astrophys. J.*, 438:763–775, 1995.

A. González-Jerez, P. Xanthopoulos, J. M. García-Regaña, I. Calvo, J. Alcusón, A. Bañón Navarro, M. Barnes, F. I. Parra, and J. Geiger. Electrostatic gyrokinetic simulations in Wendelstein 7-X geometry: benchmark between the codes stella and GENE. *J. Plasma Phys.*, 88:905880310, 2022.

T. Görler, X. Lapillonne, S. Brunner, T. Dannert, F. Jenko, F. Merz, and D. Told. The global version of the gyrokinetic turbulence code `GENE`. *J. Comp. Phys.*, 230:7053, 2011.

H Grad and H Rubin. Hydromagnetic equilibria and force-free fields. Technical report, New York Univ., New York. Inst. of Mathematical Sciences, 10 1958. URL `https://www.osti.gov/biblio/4315300`.

V. Grandgirard, M. Brunetti, P. Bertrand, N. Besse, X. Garbet, P. Ghendrih, G. Manfredi, Y. Sarazin, O. Sauter, E. Sonnendrücker, J. Vaclavik, and L. Villard. A drift-kinetic Semi-Lagrangian 4D code for ion turbulence simulation. *J. Comp. Phys.*, 217(2):395–423, 2006.

Virginie Grandgirard, Jeremie Abiteboul, Julien Bigot, Thomas Cartier-Michaud, Nicolas Crouseilles, G. Dif-Pradalier, C. Ehrlacher, D. Esteve, X. Garbet, P. Ghendrih, G. Latu, M. Mehrenberger, C. Norscini, C. Passeron, F. Rozar, Y. Sarazin, E. Sonnendrucker, A. Strugarek, and D. Zarzoso. A 5D gyrokinetic full-f global semi-Lagrangian code for flux-driven ion turbulence simulations. *Comput. Phys. Comm.*, 207:35–68, 2016.

Frederick C. Grant and Marc R. Feix. Fourier-Hermite solutions of the Vlasov equations in the linearized limit. *Phys. Fluids*, 10:696–702, 1967.

M. Greenwald, J. Terry, S. Wolfe, S. Eijima, M. Bell, S. Kaye, and G. Neilson. A new look at density limits in tokamaks. *Nucl. Fusion*, 28:2199, 1988.

Robin Greif, Frank Jenko, and Nils Thuerey. Physics-preserving AI-accelerated simulations of plasma turbulence, 2023.

Gregory W. Hammett and Francis W. Perkins. Fluid moment models for Landau damping with application to the ion-temperature-gradient instability. *Phys. Rev. Lett.*, 64:3019, 1990.

M. R. Hardman, M. Abazorius, J. Omotani, M. Barnes, S. L. Newton, J. W. S. Cook, P. E. Farrell, and F. I. Parra. A higher-order finite-element implementation of the nonlinear Fokker-Planck collision operator for charged particle collisions in a low density plasma. *Comp. Phys. Communications*, 314:109675, 2025.

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy,

Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585:357–362, 2020.

Akira Hasegawa and Kunioki Mima. Pseudo-three-dimensional turbulence in magnetized nonuniform plasma. *Phys. Fluids*, 21:87–92, 1978.

Per Helander and Dieter J. Sigmar. *Collisional Transport in Magnetized Plasmas*. Cambridge University Press, 2002.

Edmund G. Highcock, Noah R. Mandell, Michael Barnes, and William Dorland. Optimisation of confinement in a fusion reactor using a nonlinear turbulence model. *J. Plasma Phys.*, 84: 905840208, 2018.

S. P. Hirshman and J. C. Whitson. Steepest-descent moment method for three-dimensional magnetohydrodynamic equilibria. *Phys. Fluids*, 26:3553–3568, 1983.

Antoine C. D. Hoffmann, Baptiste J. Frei, and Paolo Ricci. Gyrokinetic simulations of plasma turbulence in a Z-pinch using a moment-based approach and advanced collision operators. *J. Plasma Phys.*, 89:905890214, 2023a.

Antoine C. D. Hoffmann, Baptiste J. Frei, and Paolo Ricci. Gyrokinetic moment-based simulations of the Dimits shift. *J. Plasma Phys.*, 89:905890611, 2023b.

Evans Etrue Howard, Abeyah Calpatura, Pieterjan Robbe, and Bert Debusschere. Quantification of Uncertainty in ENsembles of Data Streams (QUENDS), 2025. URL `https://github.com/sandialabs/quends`.

Gregory G. Howes, Steven C. Cowley, William Dorland, Gregory W. Hammett, Eliot Quataert, and Alexander A. Schekochihin. Astrophysical gyrokinetcs: basic equations and linear theory. *Astrophys. J.*, 651:590, 2006.

Ziyu Huang, Chuanfei Dong, and Liang Wang. Machine-learning heat flux closure for multimoment fluid modeling of nonlinear landau damping. *Proceedings Natl. Acad. Sci. U.S.A.*, 122 (11):e2419073122, 2025.

O. A. Hurricane, D. A. Callahan, D. T. Casey, A. R. Christopherson, A. L. Kritcher, O. L. Landen, S. A. Maclaren, R. Nora, P. K. Patel, J. Ralph, D. Schlossberg, P. T. Springer, C. V. Young, and A. B. Zylstra. Energy principles of scientific breakeven in an inertial fusion experiment. *Phys. Rev. Lett.*, 132:065103, 2024.

Y. Idomura, S. Tokuda, and Y. Kishimoto. Global gyrokinetic simulation of ion temperature gradient driven turbulence in plasmas using a canonical maxwellian distribution. *Nucl. Fusion*, 43: 234, 2003.

Emil R. Ingelsten, Madox C. McGrae-Menge, E. Paulo Alves, and Istvan Pusztai. Data-driven discovery of a heat flux closure for electrostatic plasma phenomena, 2024. URL `https://arxiv.org/abs/2411.18358`.

IPCC. Climate Change 2023: Synthesis Report. Contribution of Working Groups, I, II, and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Technical report, Intergovernmental Panel on Climate Change, Geneva, Switzerland, 2023.

Plamen G. Ivanov and Toby Adkins. An analytical form of the dispersion function for local linear gyrokinetics in a curved magnetic field. *J. Plasma Phys.*, 89:905890213, 2023.

H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. GMD Report 148, German National Research Center for Information Technology, 2001.

Azarakhsh Jalalvand, Alan A. Kaptanoglu, Alvin V. Garcia, Andrew O. Nelson, Joseph Abbate, Max E. Austin, Geert Verdoolaege, Steven L. Brunton, William W. Heidbrink, and Egemen Kolemen. Alfvén eigenmode classification based on ECE diagnostics at DIII-D using deep recurrrent neural networks. *Nuclear Fusion*, 62:026007, 2022.

Frank Jenko, William Dorland, and Barett N. Rogers. Electron temperature gradient driven turbulence. *Phys. Plasmas*, 7:1904–1910, 2000.

Junjie Jiang and Ying-Cheng Lai. Model-free prediction of spatiotemporal dynamical systems with recurrent neural networks: role of network spectral radius. *Phys. Rev. Research*, 1:033056, 2019.

S. Jolliet, P. Angelino A. Bottino, R. Hatzky, T. M. Tran, B. F. McMillan, O. Sauter, K. Appert, and Y. Idomuraand L. Villard. A global collisionless PIC code in magnetic coordinates. *Comp. Phys. Comm.*, 177:409–425, 2007.

Rogerio Jorge, Paolo Ricci, and Nuno F. Loureiro. A drift-kinetic analytical model for scrape-off layer plasma dynamics at arbitrary collisionality. *J. Plasma Phys.*, 86:905830606, 2017.

Glenn Joyce, Georg Knorr, and Homer K Meier. Numerical integration methods of the Vlasov equation. *J. Comput. Physics*, 8:53–63, 1971.

J. Juno, A. Hakim, J. M. TenBarge, E. Shi, and W. Dorland. Discontinuous Galerkin algorithms for fully kinetic plasmas. *J. Comp. Phys.*, 353:110–147, 2018.

Anjor Kanekar, A. A. Schekochihin, William Dorland, and Nuno F. Loureiro. Fluctuation-dissipation relations for a plasma-kinetic Langevin equation. *J. Plasma Phys.*, 81(1):305810104, 2015.

Goksin Kavlak, Magdalena M. Klemun, Ajinkya S. Kamat, Brittany L. Smith, Robert M. Margolis, and Jessika E. Trancik. Nature of innovations affecting photovoltaic system costs. *PLoS One*, 20:e0320676, 2025.

S. Kerboua-Benlarbi, R. Nouailletas, B. Faugeras, E. Nardon, and P. Moreau. Magnetic control of WEST plasmas through deep reinforcement learning. *IEEE Trans. Plasma Sci.*, 52(9):3698, 2024.

Patrick Kim, Stefan Buller, Rory Conlin, William Dorland, Daniel Dudt, Rahul Gaur, Rogerio Jorge, Egemen Kolemen, Matt Landreman, Noah Mandell, and Dario Panici. Optimization of nonlinear turbulence in stellarators. *J. Plasma Phys.*, 90:905900402, 2024.

A. N. Kolmogorov. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Proceedings of the Royal Society of London. Series A*, 434:9–13, 1991.

Robert H. Kraichnan. The structure of isotropic turbulence at very high Reynolds numbers. *J. Fluid Mech.*, 5:497–543, 1959.

Joel A. Kulesza, Terry R. Adams, Jerawan Chudoung Armstrong, Simon R. Bolding, Forrest B. Brown, Jeffrey S. Bull, Timothy Patrick Burke, Alexander Rich Clark, Robert Arthur Forster III, Jesse Frank Giron, Tristan Sumner Grieve, Colin James Josey, Roger L. Martz, Gregg W. McKinney, Eric John Pearson, Michael Evan Rising, Clell Jeffrey Solomon Jr., Sriram Swaminarayan, Travis John Trahan, Stephen Christian Wilson, and Anthony J. Zukaitis. MCNP® Code Version 6.3.0 Theory & User Manual. Technical report, Los Alamos National Laboratory, 2022.

Lev D. Landau. Die kinetische gleichung für den fall Coulombscher wechselwirkung. *Phys. Z. Sowjetunion*, 10:154–164, 1936. English translation in *Collected papers of L. D. Landau*, Pergamon Press (1965), 163-170.

Matt Landreman, Jong Youl Choi, Caio Alves, Prasanna Balaprakash, Michael Churchill, Rory Conlin, and Gareth Roberg-Clark. How does ion temperature gradient turbulence depend on magnetic geometry? insights from data and machine learning. *J. Plasma Phys.*, 91:E120, 2025.

E. Lanti, N. Ohana, N. Tronko, T. Hayward-Schneider, A. Bottino, B.F. McMillan, A. Mishchenko, A. Scheinberg, A. Biancalani, P. Angelino, S. Brunner, J. Dominski, P. Donnel, C. Gheller, R. Hatzky, A. Jocksch, S. Jolliet, Z.X. Lu, J.P. Martin Collar, I. Novikau, E. Sonnendrücker, T. Vernay, and L. Villard. `ORB5`: A global electromagnetic gyrokinetic code using the pic approach in toroidal geometry. *Comp. Phys. Comm.*, 251:107072, 2020.

J. D. Lawson. Some criteria for a useful thermonuclear reactor. *Proc. Phys. Soc. B*, 70:6, 1957.

Z. Lin, T.S. Hahm, W.W. Lee, W. M. Tang, and R. B. White. Turbulent transport reduction by zonal flows: Massively parallel simulations. *Science*, 281:1835–1837, 1998.

Nuno F. Loureiro, A. A. Schekochihin, and A. Zocco. Fast collisionless reconnection and electron heating in strongly magnetized plasmas. *Phys. Rev. Lett.*, 111(2):025002, 2013.

Nuno F. Loureiro, William Dorland, L. Fazendeiro, A. Kanekar, A. Mallet, M. S. Vilelas, and A. Zocco. `Viriato`: A Fourier-Hermite spectral code for strongly magnetized fluid-kinetic plasma dynamics. *Computer Phys. Communications*, 206:45–63, 2016.

Zhixin Lu, Jaideep Pathak, Brian Hunt, Michelle Girvan, Roger Brockett, and Edward Ott. Reservoir observers: Model-free infrerence of unmeasured variables in chaotic systems. *Chaos*, 27: 041102, 2017.

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30, pages 4765–4774. Curran Associates, Inc., 2017. URL `https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf`.

Chenhao Ma, Ben Zhu, Xue-Qiao Xu, and Weixing Wang. Machine learning surrogate models for Landau fluid closure. *Phys. Plasmas*, 27:042502, 2020.

Wolfgang Maass, Thomas Natschläger, and Henry Makram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computat.*, 14: 2531–2560, 2002.

Noah R. Mandell, William Dorland, and Matt Landreman. Laguerre-Hermite pseudo-spectral velocity formulation of gyrokinetics. *J. Plasma Phys.*, 84(1):905840108, 2018.

Noah R. Mandell, William Dorland, Ian Abel, Rahul Gaur, Patrick Kim, Mike Martin, and Tony Qian. GX: a GPU-native gyrokinetic turbulence code for tokamak and stellarator design. *J. Plasma Phys.*, 90:905900402, 2024.

Bertrand Mercier. *An Introduction to the Numerical Analysis of Spectral Methods*, chapter 1, pages 1–5. Springer Berlin, 1989.

Romain Meyrand, Anjor Kanekar, William Dorland, and Alexander A. Schekochihin. Fluidization of collisionless plasma turbulence. *Proceedings Natl. Acad. Sci. U.S.A.*, 116(4):1185–1194, 2019.

R. L. Miller, M. S. Chu, J. M. Greene, Y. R. Lin-Liu, and R. E. Waltz. Noncircular, finite aspect ratio, local equilibrium model. *Phys. Plasmas*, 5:973–978, 1998.

W. M. Nevins, G. W. Hammett, A. M. Dimits, W. Dorland, and D. E. Shumaker. Discrete particle noise in particle-in-cell simulations of plasma microturbulence. *Phys. Plasmas*, 12:122305, 2005.

Dwight R. Nicholson. *Introduction to Plasma Theory*, chapter 6. John Wiley & Sons, 1983.

R. Nies, F. Parra, M. Barnes, N. Mandell, and W. Dorland. Saturation of magnetized plasma turbulence by propagating zonal flows. *Phys. Rev. Research*, 8:013295, 2026.

NIST. The NIST reference on constants, units, and uncertainty, 2024. https://physics.nist.gov/cgi-bin/cuu/Value?mpsme.

Thomas O'Neil. Collisionless damping of nonlinear plasma oscillations. *Phys. Fluids*, 8: 2255–2262, 1965.

Joseph T. Parker and Paul J. Dellar. Fourier-Hermite spectral representation for the vlasov-poisson system in the weakly collisional limit. *J. Plasma Phys.*, 81(2):305810203, 2015.

Jaideep Pathak, Brian Hunt, Michelle Girvan, Zhixin Lu, and Edward Ott. Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach. *Phys. Rev. Lett.*, 120:024102, 2018.

Andrea Piccione, Jack W. Berkery, Steven A. Sabbagh, and Yiannis Andreopoulos. Physics-guided machine learning approaches to predict the ideal stability properties of fusion plasmas. *Nuclear Fusion*, 60:046033, 2020.

Gabriel G. Plunk, Steven C. Cowley, Alexander A. Schekochihin, and Tomo Tatsuno. Two-dimensional gyrokinetic turbulence. *J. Fluid Mech.*, 664:407–435, 2010.

Ryan Pyle, Nikola Jovanovic, Devika Subramanian, Krishna V. Palem, and Ankit B. Patel. Domain-driven models yield better predictions at lower cost than reservoir computers in Lorenz systems. *Phil. Trans. R. Soc. A*, 379:20200246, 2021.

Yilan Qin, Jiayu Ma, Mingle Jiang, Chuanfei Dong, Haiyang Fu, Wenjie Cheng, and Yaqiu jin. Data-driven modeling of Landau damping by physics-informed neural networks. *Phys. Rev. Research*, 5:033079, 2023.

Cristina Rea, Kevin Joseph Montes, Keith G. Erickson, Robert S. Granetz, and Roy Alexander Tinguely. A real-time machine learning-based disruption predictor on DIII-D. *Nuclear Fusion*, 59:096016, 2019.

Frederick Reif. *Fundamentals of statistical and thermal physics*. McGraw-Hill, 1965.

Osborne Reynolds. On the dynamical theory of incompressible viscous fluids and the determination of the criterion. *Philosophical Transactions of the Royal Society of London. (A.)*, 186: 123–164, 1895.

Gareth T. Roberg-Clark, G. G. Plunk, P. Xanthopoulos, C. Nührenberg, S. A. Henneberg, and H. M. Smith. Critical gradient turbulence optimization toward a compact stellarator reactor concept. *Phys. Rev. Research*, 5:L032030, 2023.

Pablo Rodriguez-Fernandez, Nathan T. Howard, and Jeff Candy. Nonlinear gyrokinetic predictions of SPARC burning plasma profiles enabled by surrogate modeling. *Nuclear Fusion*, 62:076036, 2022.

Paul K. Romano, Nicholas E. Horelik, Bryan R. Herman, Adam G. Nelson, Benoit Forget, and Kord Smith. OpenMC: A state-of-the-art monte carlo code for research and development. *Ann. Nucl. Energy*, 82:90–97, 2015.

Marshall N. Rosenbluth, William M. MacDonald, and David L. Judd. Fokker-Planck equation for an inverse-square force. *Phys. Rev.*, 107, 1957.

Jaemin Seo, SangKyeun Kim, Azarakhsh Jalalvand, Rory Conlin, Andrew Rothstein, Joseph Abbate, Keith Erickson, Josiah Wai, Ricardo Shousha, and Egemen Kolemen. Avoiding fusion plasma tearing instability with deep reinforcement learning. *Nature*, 626:746–751, 2024.

S. Shiraiwa, O. Meneghini, R. Parker, P. Bonoli, M. Garrett, M. C. Kaufman, J. C. Wright, and S. Wukitch. Plasma wave simulation based on a versatile finite element solver. *Phys. Plasmas*, 17:056119, 2010.

Vaclav Smil. *Energy and Civilization: A History*, chapter 5. The MIT Press, 2017a.

Vaclav Smil. *Energy and Civilization: A History*, chapter 6. The MIT Press, 2017b.

Elias M. Stein and Rami Shakarchi. *Fourier Analysis: An Introduction*, page 79. Princeton University Press, 2003.

S. Sudo, Y. Takeiri, H. Zushi, F. Sano, K. Itoh, K. Kondo, and A. Iiyoshi. Scalings of energy confinement and density limit in stellarator/heliotron devices. *Nucl. Fusion*, 30:11, 1990.

Bogdan Teaca, Frank Jenko, and Daniel Told. Gyrokinetic turbulence: between idealized estimates and a detailed analysis of nonlinear energy transfers. *New J. Phys.*, 19:045001, 2017.

Abhishek Tiwari, Joydeep Das, Jaya Kumar Alageshan, Gareth Roberg-Clark, Gabriel Plunk, Pavlos Xanthopoulos, Sarveshwar Sharma, Zhihong Lin, and Animesh Kuley. Zonal flow suppression of turbulent transport in the optimized stellarators w7-x and qstk. *Plasma Phys. and Controlled Fusion*, 67(8):085025, 2025. doi: 10.1088/1361-6587/adf91b.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

P. R. Vlachas, J. Pathak, B. R. Hunt, T. P. Sapsis, M. Girvan, E. Ott, and P. Koumoutsakos. Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics. *Neural Networks*, 126:191–217, 2020.

A. A. Vlasov. The vibrational properties of an electron gas. *Sov. Phys. Usp.*, 10(6):721, 1968.

L. Wang, A. Hakim, J. Ng, and K. Germaschewski. Exact and locally implicit source term solvers for multifluid-Maxwell systems. *J. Comp. Phys.*, 415:109510, 2020.

W. X. Wang, Z. Lin, W. M. Tang, W. W. Lee, S. Ethier, J. L. V. Lewandowski, G. Rewoldt, T. S. Hahm, and J. Manickam. Gyro-kinetic simulation of global turbulent transport properties in tokamak experiments. *Phys. Plasmas*, 3:092505, 2006.

T. H. Watanabe and H. Sugama. Velocity–space structures of distribution function in toroidal ion temperature gradient turbulence. *Nucl. Fusion*, 46:24, 2006.

Anne E. White. Validation of nonlinear gyrokinetic transport models using turbulence measurements. *J. Plasma Phys.*, 85:925850102, 2019.

Samuel E. Wurzel and Scott C. Hsu. Progress toward fusion energy breakeven and gain as measured against the Lawson criterion. *Phys. Plasmas*, 29:062103, 2022.

M. Yoshida, R. M. McDermott, C. Angioni, Y. Camenen, J. Citrin, M. Jakubowski, J. W. Hughes, Y. Idomura, P. Mantica, A. Mariani, S. Mordijck, E. J. Paul, T. Tala, G. Verdoolaege, A. Zocco, F. J. Casson, G. Dif-Pradaliler, B. Duval, B. A. Grierson, S. M. Kaye, P. Manas, M. Maslov, T. Odstrcil, J. E. Rice, L. Schmitz, F. Sciortino, E. R. Solano, G. Staebler, M. Valovič, E. Wolfrum, J. A. Snipes, and the Transport and Confinement Group. Transport and confinement physics: Chapter 2 of the special issue: on the path to tokamak burning plasma operation. *Nuclear Fusion*, 65:033001, 2025.

A. Zocco and A. A. Schekochihin. Reduced fluid-kinetic equations for low-frequency dynamics, magnetic reconnection, and electron heating in low-beta plasmas. *Phys. Plasmas*, 18(10): 102309, 2011.