UNIVERSITY OF CALIFORNIA SAN DIEGO

Topics in Mesoscopic Turbulent Transport

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Physics with a Specialization in Computational Science

by

Robin A. Heinonen

Committee in charge:

Professor Patrick H. Diamond, Chair
Professor Daniel P. Arovas
Professor Stefan G. Llewellyn-Smith
Professor George R. Tynan
Professor Yi-Zhuang You

2021

The dissertation of Robin A. Heinonen is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2021

DEDICATION

I dedicate this thesis to my father, Olle Heinonen, who is a brilliant physicist, a patient mentor, my role model, my friend, and the best man I know.

EPIGRAPH

[Q]uotation [. . .] is a serviceable substitute for wit.

—W. Somerset Maugham

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Thanks first to my advisor for the last five and a half years, Pat Diamond, who is a physicist of the highest caliber with unparalleled knowledge and understanding of plasma, fluid, and nonlinear physics. I thank him for his patience with me as I have fumbled my way around this challenging field, his willingness to entertain my naïve research ideas, and his tolerance of my unusual sleep habits. I am grateful for all the support he has given me over the years, whether explaining a concept to me for the dozenth time or writing letters on my behalf for conferences, fellowships, and other opportunities. Pat's work ethic and standards of excellence are inspiring, and I thank him for always pushing me to do the highest quality work of which I am capable and to never settle for mediocrity.

I also give thanks to my lovely wife, Monica Wright, who has been my rock since I met her just one-and-a-half years ago. Without her, life during the present pandemic would be lonely and miserable, but instead it has been full of joy and love. I am excited for our coming adventures together.

Thanks in addition to my wonderful parents, Olle Heinonen and Leslie McDonald, whose infinite love and support I've depended on both during the course of my Ph.D. and in the many years leading up to it. They have shaped me into the person I am today, and I owe to them all of my best qualities.

There are countless other people in my life over the past few years who have made the past few years bearable — too many to name. However, I would be remiss if I did not mention the rest of my beloved family, including my brother Nils, my sister Lindsay, and my grandmother JoAnn Magistad; my friends, both old (especially my enduring network of friends from college) and new (Jonathan Lam, Saurabh Mogre, Shruti Singh, Clayton Anderson, Norman Cao, Samantha Chen, Becky Masline, and many other graduate students, as well as Jason Cheng and the rest of the quizbowl team); faculty and staff at UCSD who have given me mentorship and support, including Sharmila Poddar, John McGreevy, Dan Arovas, and the rest of my committee; my mentors and

colleagues elsewhere: at Los Alamos (Didier Saumon and Charlie Starrett), in France (Massimo Vergassola, Guilhem Dif-Pradalier, Xavier Garbet, David Zarzoso, and Yanick Sarazin), and in the UK (Steve Tobias); and (last but certainly not least) my cat, Maya.

Chapter 2, in full, is a reprint of the material as it appears in R. A. Heinonen and P. H. Diamond, *Plasma Phys. Control. Fusion* 62, 105017 (2020). The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in R. A. Heinonen and P. H. Diamond, *Physics of Plasmas* 27, 032303 (2020). The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is currently being prepared for submission for publication of the material. R. A. Heinonen and P. H. Diamond (2021). The dissertation author was the primary investigator and author of this material.

VITA

2014	B. S. in Mathematics *magna cum laude* and B. S. in Physics, University of Minnesota Twin Cities
2014	Computational Physics Workshop Fellow, Los Alamos National Laboratory
2015	Post-Baccalaureate Intern, Los Alamos National Laboratory
2021	Ph.D. in Physics with a Specialization in Computational Science, University of California San Diego

PUBLICATIONS

R. A. Heinonen and P. H. Diamond. “Learning how structures form in drift-wave turbulence." *Plasma Phys. Control. Fusion* 62, 105017 (2020).

R. A. Heinonen and P. H. Diamond. “Turbulence model reduction by deep learning.” *Phys. Rev. E* 101, 061201(R) (2020).

R. A. Heinonen, D. Saumon, J. Daligault, C. E. Starrett, S. D. Baalrud, and G. Fontaine. “Diffusion coefficients in the envelopes of white dwarfs.” *Astrophys. J.* 896.1, 2 (2020).

R. A. Heinonen and P. H. Diamond. “A closer look at turbulence spreading: How bistability admits intermittent, propagating turbulence fronts.” *Phys. Plasmas* 27, 032303 (2020).

R. A. Heinonen and P. H. Diamond. “Subcritical turbulence spreading and avalanche birth.” *Phys. Plasmas* 26, 030701 (2019).

N. M. Gill, R. A. Heinonen, C. E. Starrett, and D. Saumon. “Ion-ion dynamic structure factor of warm dense mixtures.” *Phys. Rev. E* 91, 063109 (2015).

R. A. Heinonen, E. G. Kalnins, W. Miller, Jr., and E. Subag. “Structure relations and Darboux contractions for 2D 2nd order superintegrable systems.” *SIGMA* 11, 043 (2015).

ABSTRACT OF THE DISSERTATION

Topics in Mesoscopic Turbulent Transport

by

Robin A. Heinonen

Doctor of Philosophy in Physics with a Specialization in Computational Science

University of California San Diego, 2021

Professor Patrick H. Diamond, Chair

The interaction of turbulent microscales in a plasma can conspire to produce transport on mesoscopic or macroscopic scales, often with extremely important consequences for the system. In this dissertation, we present research on mesoscopic turbulent transport in plasmas relevant to fusion energy and astrophysics.

The research is divided into three projects. First, we study drift-wave turbulence, a paradigm for gradient-driven turbulence in the tokamak, using a novel data-driven method. Deep learning is used to infer, from direct numerical simulation of the 2-D Hasegawa-Wakatani system, the dependencies of the turbulent fluxes which control the nonlinear dynamics on mesoscales, thus closing a simplified 1-D mean-field model. Using this approach, we show that the gradient of vorticity drives a nondiffusive particle flux. This nondiffusive flux, which we also recover analytically, modulates the profile in the presence of a quasiperiodic zonal flow. We also show that zonal flow formation is described by a Cahn-Hilliard-type model, both concurring with and expanding upon previous theoretical work.

The self-propagation of turbulence, called “turbulence spreading,” can lead to turbulence invading linearly stable regions of a tokamak. In the second project, we introduce a new model for turbulence spreading based on turbulence bistability. This model takes the form a reaction-diffusion equation with cubic nonlinearity and nonlinear diffusion. We find that the bistable model remedies certain deficiencies of the popular, unistable Fisher model, for instance predicting far more robust penetration of turbulence into stable regions. We also find that the model exhibits a threshold for propagation of an initial seed of turbulence, which we liken to an avalanche. We analytically estimate this threshold using a simple physical argument.

Finally, we study momentum transport in magnetohydrodynamic (MHD) turbulence on a β-plane, which serves as a simple model for the solar tachocline. We show that the cross-helicity, which is conserved in pure MHD turbulence, builds up to a predictable level in this system. Using weak turbulence theory, we also show that the cross helicity spectrum is equivalent to the Elsässer alignment spectrum, which determines momentum transport. We supplement and verify our results with direct numerical simulations.

Chapter 1

Introduction

1.1 Confinement problem

The goal of the fusion energy program is to sustain a nuclear fusion interaction long enough to produce useful electricity. Many of the challenges of this problem can be summarized by the Lawson criterion [Law57] (here expressed for the deuterium-tritium reaction):

$$n_e T \tau_E \geq 3 \times 10^{21} \text{ keV s/m}^3. \tag{1.1}$$

That is, the electron density, temperature, and *confinement time* τ_E must all be sufficiently large if the fusion reaction is to produce more energy than is externally supplied. The confinement time measures the inverse rate of energy loss from the plasma (for fusion reactions generally require plasma conditions) to the environment. Adequately confining the plasma while sustaining the requisite plasma pressure for fusion has proved to be a highly nontrivial problem.

Along with inertial confinement, which is not pertinent to this dissertation, *magnetic confinement* is one of the central paradigms for confining a fusion plasma. There are a number of practical designs which employ a magnetic field to confine a plasma, but among the most enduring and well-studied is the toroidally-shaped device called the *tokamak*. Because charged particles will tend to move along magnetic field lines (while gyrating in the perpendicular plane with the Larmor frequency $\omega_c = qB/mc$), a natural solution to the confinement problem might seem to be to "glue" the ends of a solenoid together into a torus. However, for geometric reasons, the toroidal field in such a configuration is not uniform, which leads the plasma to experience radial motion according to the "grad-B drift"

$$\mathbf{v}_{\nabla B} = \frac{E_\perp}{qB} \frac{\mathbf{B} \times \nabla B}{B^2}. \tag{1.2}$$

Other kinds of inhomogeneities can give rise to a somewhat expansive zoology of additional radial drifts, all of which threaten to destroy confinement. The tokamak's solution is twist the toroidal magnetic field lines into helices, which, on average, cancels the inward and outward radial drift motion. The physics of the motion of charged particles in a tokamak configuration is quite rich but is not the focus of this dissertation.

Though the magnetic drifts are resolved for the moment, the tokamak nevertheless experiences a host of difficult challenges due to the development of turbulence and turbulent transport, which results due to instabilities set up by temperature and density gradients in the device. One universal instability is the *drift-wave* instability, which is a useful paradigm which describes a number of turbulence phenomena in the tokamak. We now turn our attention to drift waves.

1.2 Drift-wave turbulence

1.2.1 Phenomenology of drift waves

Radial density and temperature gradients (and thus pressure gradients) are generic to tokamak devices, as typically heat and particle sources are located somewhere in the core. The pressure gradients, in tandem with the toroidal field, set up radial diamagnetic drifts which depend on species α (ion or electron)

$$v_{d,\alpha} = \frac{1}{q_\alpha n_\alpha B} \frac{\partial p_\alpha}{\partial r}. \tag{1.3}$$

In the presence of a localized charge separation, there will be an additional $E \times B$ drift

$$\mathbf{v}_E = -\frac{c}{B^2} \nabla \tilde{\phi} \times \mathbf{B}, \tag{1.4}$$

where $\tilde{\phi}$ is the electric potential fluctuation induced by the charge separation, in order to satisfy the force balance

$$q(-\nabla \tilde{\phi} + \frac{1}{c} \mathbf{v} \times \mathbf{B}) = \nabla p / n_0. \tag{1.5}$$

Note that the $E \times B$ drift does *not* depend on the species of the charge.

Drift waves are convective cells, with frequency $\omega \simeq \omega_{*,\alpha} = \mathbf{k} \cdot \mathbf{v}_{d,\alpha}$, which circulate about a localized charge separation at the $E \times B$ drift velocity [Hor99].

Consider a fluctuation $\tilde{n}_i$ from the background density n_0. There will be an associated electron density fluctuation $\tilde{n}_e \simeq \tilde{n}_i$, which is the principle of quasineutrality. To see this, Poisson's equation gives the associated fluctuation in electric potential:

$$\nabla^2 \tilde{\phi} = -4\pi e(\tilde{n}_i - \tilde{n}_e), \tag{1.6}$$

which gives

$$\frac{\tilde{n}_i - \tilde{n}_e}{n_0} \simeq \frac{\lambda_D^2}{\ell^2} \frac{e\tilde{\phi}}{T_e} \simeq 0 \tag{1.7}$$

on scales ℓ larger than the electron Debye length $\lambda_D = (T_e / 4\pi e^2 n_0)^{1/2}$. By assuming the parallel electron dynamics are fast compared to ω_* ("adiabatic"), we can argue that the electrons will set up a Boltzmann distribution

$$n_e = n_0 \exp(e\tilde{\phi}/T_e), \tag{1.8}$$

according to the parallel force balance $n_e e \partial_z \tilde{\phi} = \partial_z (n_e T_e) \simeq T_e \partial_z n_e$. Expanding the exponential, we have

$$\frac{\tilde{n}_e}{n_0} \simeq \frac{e\tilde{\phi}}{T_e}. \tag{1.9}$$

Thus, adiabatic electron density fluctuations will be tightly coupled to potential fluctuations. This relationship, however, breaks down in a real system on account of a number of effects, including particle collisions and wave-particle resonances. The precise relationship is complicated and depends on a number of details of the system, but in general, the nonadiabatic behavior can be summarized by the development of a phase shift between the density and potential $\delta_{\mathbf{k}}$:

$$\tilde{n}_{e,\mathbf{k}} \simeq \frac{e\tilde{\phi}_{\mathbf{k}}}{T_e}(1 - i\delta_{\mathbf{k}}). \tag{1.10}$$

Provided $\delta_{\mathbf{k}} > 0$, this phase shift will lead to exponential growth in time of the drift-wave amplitude. When many drift modes thusly become unstable and interact, complicated nonlinear phenomena dominate the dynamics, and the system becomes *turbulent*. In the next subsection, we formalize this idea by presenting the simplest realistic model for drift-wave turbulence, the Hasegawa-Wakatani model.

1.2.2 Hasegawa-Wakatani model

The Hasegawa-Wakatani (HW) model [HW83, WH84] is a model for *collisional* drift-wave turbulence in response to an electron density gradient. It serves as a useful theoretical paradigm for studying turbulence and transport in the tokamak.

The starting point to derive this system is to assume cold ions $T_i \ll T_e$, so that $\nabla p_i = 0$, and a strong toroidal field $\mathbf{B} = B\hat{z}$. This gives the ion momentum equation

$$\partial_t \mathbf{v}_i + \mathbf{v}_i \cdot \nabla \mathbf{v}_i = \frac{e}{m_i}\left(-\nabla\phi + \frac{1}{c}\mathbf{v}_i \times \mathbf{B}\right) + \nu\nabla^2\mathbf{v}_i, \tag{1.11}$$

where ν is the viscosity. At first order in $1/\omega_{c,i}\partial_t$, equilibrium is attained when

$$\mathbf{v}_i \simeq \mathbf{v}_E - \frac{c}{\omega_{c,i}B}\frac{d\nabla\phi}{dt} + \frac{\nu c}{\omega_{c,i}B}\nabla^2(\nabla\phi), \tag{1.12}$$

where $\omega_{c,i}$ is the ion Larmor radius, $\mathbf{v}_E = -\frac{c}{B}\nabla\phi\times\hat{z}$ is the $E\times B$ drift and $d/dt = \partial_t + \mathbf{v}_E\cdot\nabla$. The second term is called the *polarization drift* and allows for compression of ions. It must be kept because $\nabla\cdot\mathbf{v}_E = 0$; without it, quasineutrality could not be satisfied. The third term is an effect due to ion collisions.

This expression is now substituted into the ion continuity equation

$$\partial_t n + \nabla\cdot(n\mathbf{v}_i) = 0, \tag{1.13}$$

where we've invoked quasineutrality to write $n_e = n_i = n$. One thus obtains

$$\frac{d}{dt}\left(\frac{c}{\omega_{c,i}B}\nabla^2\phi - \log n\right) = \frac{\nu c}{\omega_{c,i}B}\nabla^4\phi. \tag{1.14}$$

For the electrons, we ignore the polarization drift because the mass ratio m_e/m_i is small, and write $\mathbf{v}_e = \mathbf{v}_E - \frac{1}{ne}J_{e,\|}$. The parallel electron current is determined by Ohm's law

$$\eta J_{e,\|} = -\partial_z\phi + \frac{1}{ne}\partial_z(nT_e) \simeq \frac{T_e}{e}\partial_z\left(\frac{e\phi}{T_e} - \log n\right), \tag{1.15}$$

where η is the parallel resistivity. Inserting this into the electron continuity equation, we have

$$\frac{d\log n}{dt} = \frac{T_e}{e^2\eta}\partial_z\left[\frac{1}{n}\partial_z\left(\log n - \frac{e\phi}{T_e}\right)\right] \simeq \frac{T_e}{n_0e^2\eta}\partial_z^2\left(\log n - \frac{e\phi}{T_e}\right). \tag{1.16}$$

Finally, we have the HW equations

$$\frac{d\log n}{dt} = \frac{T_e}{n_0 e^2 \eta}\partial_z^2 \left(\log n - \frac{e\phi}{T_e}\right) \tag{1.17}$$

$$\frac{d}{dt}\nabla^2\phi = \frac{\omega_{c,i} B T_e}{n_0 c e^2 \eta}\partial_z^2 \left(\log n - \frac{e\phi}{T_e}\right) + \nu\nabla^4\phi. \tag{1.18}$$

It is standard to choose the normalizations $\log n \to n$, $e\phi/T_e \to \phi$, $x/\rho_s \to x$, $y/\rho_s \to y$, $\omega_{c,i}t \to t$, and $\nu/\rho_s^2\omega_{c,i} \to \nu$, where $\rho_s = \sqrt{\frac{T_e}{m_i}}\omega_{c,i}^{-1}$. This simplifies the system to

$$\frac{dn}{dt} = \alpha\left(\phi - n\right) \tag{1.19}$$

$$\frac{d}{dt}\nabla^2\phi = \alpha\left(\phi - n\right) + \nu\nabla^4\phi, \tag{1.20}$$

where $\alpha = -\frac{T_e}{n_0 e^2 \eta\, \omega_{c,i}}\partial_z^2$ is called the *adiabatic operator*, and now $d/dt = \partial_t + (\hat{z}\times\nabla\phi)\cdot\nabla$.

Let us discuss the physics of this system. First, it is easy to show that the system conserves two independent quadratic invariants up to dissipation: the energy

$$E = \int d^3\mathbf{x}\,(n^2 + (\nabla\phi)^2) \tag{1.21}$$

and the *potential enstrophy*

$$W = \int d^3\mathbf{x}\,(n - \nabla^2\phi)^2. \tag{1.22}$$

. The system also conserves the *potential vorticity* $\mathrm{PV} = n - \nabla^2\phi$.

The electric potential serves as a streamfunction for the flow, $\mathbf{v} = \hat{z}\times\nabla\phi$, so a local charge separation $\nabla^2\phi$ in this system is equivalent to a local vorticity.

It is useful to consider the limits $\alpha \to \infty$ and $\alpha \to 0$, where here α is considered to be a c-number, with $\partial_z^2 \to -k_z^2$. In the former case, the system rapidly drives to a manifold on which $\tilde{\phi} = \tilde{n}$, so the electrons are Boltzmann (hence the name, "adiabatic"). Then, if we assume the background density varies only in the $\hat{x}$ direction, $n = \kappa x + \tilde{n}(\mathbf{x})$, the system collapses to the single equation

$$\frac{d}{dt}(\nabla^2\phi - \phi) - \kappa\partial_y\phi = \nu\nabla^4\phi, \tag{1.23}$$

which expresses conservation of PV. This *Hasegawa-Mima equation* [HM77] does not contain an instability and so must be externally forced in order to exhibit turbulence. In addition to plasma physics, it is used in geophysics as a simple model of atmospheric flows [Cha71], where the excitations are known as Rossby waves rather than drift waves. In the absence of a mean flow, the linear modes of this equation have $\omega = k_y\kappa/(1+k^2)$. This, at last, is the drift wave frequency, which is inversely proportional to the density gradient scale length $1/L_n$. Note that there is a family of zero-frequency modes with $k_y = 0$. These *zonal modes* can be viewed as arbitrary mean flows $V(x)$ and are of great importance to drift-wave turbulence, and will be discussed in depth in the following subsection.

In the hydrodynamic limit, we obtain the 2-D Navier-Stokes system, which describes a neutral fluid flow on a plane. This system is of significant interest in its own right, but we will not study it in depth here.

Let us now derive the dispersion relation for drift waves for the whole HW system. Assume there is no background flow and that the background density varies only in the $\hat{x}$ direction, $n = \kappa x + \tilde{n}(\mathbf{x})$. Then the density equation becomes

$$\partial_t\tilde{n} + (\hat{z}\times\nabla\phi)\cdot\nabla\tilde{n} + \kappa\partial_y\phi = \alpha(\tilde{n} - \phi). \tag{1.24}$$

Then, if we assume a perturbation of the form $\tilde{n} = \tilde{n}_{\mathbf{k},\omega}e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$, $\phi = \tilde{\phi}_{\mathbf{k},\omega}e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$, insert into the HW system, and linearize, we obtain (ignoring dissipation)

$$-i\omega\tilde{n}_{\mathbf{k},\omega} + ik_y\kappa\tilde{\phi}_{\mathbf{k},\omega} = \alpha(\tilde{\phi}_{\mathbf{k},\omega} - \tilde{n}_{\mathbf{k},\omega}) \tag{1.25}$$

$$ik^2\omega\tilde{\phi}_{\mathbf{k},\omega} = \alpha(\tilde{\phi}_{\mathbf{k},\omega} - \tilde{n}_{\mathbf{k},\omega}), \tag{1.26}$$

The system has nontrivial solutions if the dispersion relation

$$k^2\omega^2 + i\alpha(1+k^2)\omega - ik_y\kappa\alpha = 0 \tag{1.27}$$

is satisfied. In the *adiabatic limit* $\alpha \to \infty$, the solution is

$$\omega_r \simeq \frac{k_y\kappa}{1+k^2} \tag{1.28}$$

$$\gamma \simeq \frac{k_y^2\kappa^2}{\alpha(1+k^2)^3}, \tag{1.29}$$

where we have written $\omega = \omega_r + i\gamma$. Hence, in the presence of any background density gradient, the drift-wave amplitude will grow exponentially in time. In practice (that is, in the presence of damping effects), there will be a finite critical gradient κ_c such that we must have $|\kappa| > \kappa_c$ if unstable modes are to be supported.

When the drift modes have reached sufficient amplitude, the (convective) nonlinear terms are important and the system becomes turbulent. It is well-known that the Hasegawa-Wakatani system exhibits a *dual cascade*: the energy goes to large scales (i.e. small wavenumber), whereas the enstrophy, that is, the mean-squared vorticity, goes to small scales (large wavenumber), where it is dissipated.

1.2.3 Zonal flows

Any discussion of drift-wave turbulence would be incomplete without addressing the zonal modes, hereafter referred to as *zonal flows*, which are axisymmetric, zero-frequency modes with zero poloidal and toroidal wavenumber [DIIH05a]. In the adiabatic regime of HW, the inverse cascade is realized by the spontaneous formation of large, quasiperiodic, and radially sheared zonal flows. As will be discussed in the following section, this zonal flow formation is the result of *nonlocal* (in wavenumber space) interactions between unstable drift-wave modes.

Zonal flows are of extreme interest to the confinement problem. Interest in tokamaks was revived in 1982 with the discovery of the *high-confinement mode* [WBB$^+$82] or H-mode: when the heating power is increased beyond a certain threshold, the device spontaneously transitions to a regime with much better confinement properties. The H-mode is characterized by the presence of turbulence-generated, radially sheared zonal flows, which are now understood to be the origin of the H-mode's desirable properties.

Zonal flows are considered benign repositories of turbulence energy because they cannot drive a radial $E \times B$ flow and thus do not cause harmful transport. Moreover, they are robust in the sense that, due to their symmetry properties, they are (a) modes of *minimal inertia*, since they are not screened by Boltzmann electrons (unlike other drift modes), and thus can attain large speeds, and (b) do not experience nonlinear Landau damping, and thus are only regulated by collisional damping.

Finally, zonal flows are understood to regulate turbulence by "eddy shearing." Because zonal flows are radially sheared, a turbulent eddy will gradually be distorted and stretched in the poloidal direction by a zonal flow until it is destroyed. At the level of eikonal theory (i.e the Hamiltonian ray equations of geometric optics) in the presence of a mean flow shear V_y', the radial wavenumber of a drift mode evolves as

$$k_x = k_{x,0} - k_y V_y' t. \tag{1.30}$$

The second term is the result of the inhomogeneous Doppler shift induced by the mean flow. Thus, the radial wavenumber increases, radially elongating the eddy and enhancing its coupling to small-scale dissipation.

It is often useful to view drift wave turbulence and zonal flows as separate populations which interact via a "predator-prey" feedback loop [DLCT94]: the drift waves (the prey) grow due to the gradient instability drive, while the zonal flows (the predators) "feed" upon the drift

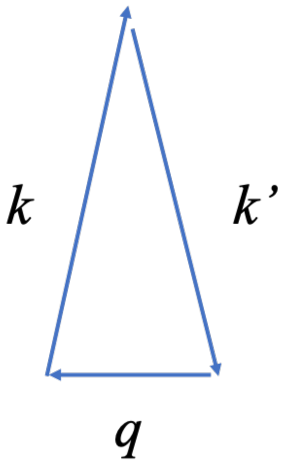


Figure 1.1: A multiscale interaction between two small scales **k** and **k**′ and a large scale **q**.

wave population and are themselves quenched by collisional processes. This can be summarized by a zero-dimensional toy model for the drift wave energy N and the zonal flow energy Z of the form

$$\partial_t N = \alpha N - \beta NZ - \gamma N^2 \tag{1.31}$$

$$\partial_t Z = \beta NZ - \delta Z - \varepsilon Z^2. \tag{1.32}$$

Here, α is the linear turbulence drive, β is the nonlinear zonal flow drive (note that this interaction conserves energy), δ is linear flow damping, and γ and ε represent nonlinear damping processes.

1.3 How meso- and macro-scale transport emerges from small-scale interaction

A unifying theme of this dissertation is that multiscale interactions can lead to meso- or macro-scale turbulent transport. (Here, “macro” means on the order of the system size, and “meso” means larger than the turbulent microscales but smaller than the system size.) For a triplet interaction, such a multiscale interaction amounts to two turbulent, microscale modes $\mathbf{k}$ and $\mathbf{k}'$ which are nearly antiparallel, $\mathbf{k} \simeq -\mathbf{k}'$, beating to form a third mode with wavevector $\mathbf{q}$, where $q \ll k, k'$. This can be visualized by a thin isosceles triangle in wavevector space (see Fig. 1.1).

This sort of *nonlocal* interaction between disparate scales is extremely important in plasma turbulence, and it diverges strongly from the usual cascade picture, where one imagines energy being transferred smoothly from scale to slightly smaller scale until it reaches dissipation.

As an especially simple example, the interaction of unstable drift modes can produce a turbulent particle flux with characteristic scale length $L_n \gg \ell_{\mathrm{turb}} \simeq \rho_s$. This can be shown using the methods of quasilinear theory [VVS61, VVS62], which approximate the nonlinear dynamics by the interaction of coherent linear modes. The radial particle flux generated by drift modes $\mathbf{k}$ and $\mathbf{k}' = \mathbf{q} - \mathbf{k}$, with $q \ll k$, is

$$\Gamma_q = \tilde{v}_{x,\mathbf{k}} \tilde{n}_{\mathbf{k}'} = \mathrm{Re}\, i k_y \tilde{\phi}_{\mathbf{k}} \tilde{n}_{\mathbf{k}'} \simeq -\frac{1}{\alpha} \frac{k^2 k_y^2}{1+k^2} \frac{d\langle n \rangle}{dx} |\tilde{\phi}_{\mathbf{k}}|^2, \tag{1.33}$$

which varies over the gradient scale length — a meso-scale. Note that this turbulent particle flux is *diffusive*, proportional to the negative density gradient, with a nonlinear diffusivity which scales with the turbulent intensity. Thus, the turbulence tends to relax the gradient which drives it via linear instability.

We now turn our attention to two more complex kinds of multiscale interactions which are especially pertinent to this dissertation: the nonlinear production of zonal flows, and the spreading of turbulence.

1.3.1 Zonal flow production

Zonal flow production (crudely) is a special case of Fig. 1.1 where the resultant wavevector is purely radial, $\mathbf{q} = q\hat{r}$. Zonal flows are generated via the turbulent Reynolds stress $\langle \tilde{v}_x \tilde{v}_y \rangle$, which is the radial flux of poloidal momentum. This can be seen by separating the electric potential into zonal (i.e. poloidally and toroidally symmetric) and nonzonal components, $\phi = \langle\phi\rangle + \tilde{\phi}$ and substituting into the HW momentum equation, yielding

$$\partial_t V_y = \partial_x \langle \partial_x \tilde{\phi} \partial_y \tilde{\phi} \rangle - \mu V_y + \nu \partial_x^2 V_y, \tag{1.34}$$

where $V_y = \partial_x \langle\phi\rangle$ is the mean poloidal flow, and μ is a linear damping term. This was obtained by applying the Taylor identity (see Appendix A) and "uncurling" the vorticity equation.

Consider the interaction of four modes: a primary drift mode $\mathbf{p}$, a zonal mode $\mathbf{q} = q\hat{x}$, and two sidebands $\mathbf{p}_\pm = \mathbf{p} \pm \mathbf{q}$. At the level of the (collisionless) HM equation, Fourier modes evolve according to

$$\partial_t \phi_{\mathbf{k}} + i\omega_{\mathbf{k}} \phi_{\mathbf{k}} = \sum_{\mathbf{k}'} \frac{k'^2}{1+k^2} \hat{z} \cdot (\mathbf{k} \times \mathbf{k}') \phi_{\mathbf{k}'} \phi_{\mathbf{k}-\mathbf{k}'}, \tag{1.35}$$

where $\omega_{\mathbf{k}} = k_y \kappa / (1+k^2)$. Here, we consider the sum to be just over the truncated system of four modes. Now, write $\phi_{\mathbf{k}} = \hat{\phi}_{\mathbf{k}} e^{-i\omega_{\mathbf{k}} t}$ for each $\mathbf{k}$, assume that the primary mode amplitude is large, and linearize these equations with respect to the other three modes to obtain the system

$$\partial_t \hat{\phi}_{\mathbf{q}} = \frac{p_y q}{1+q^2} \left((p_+^2 - p^2) \hat{\phi}_{\mathbf{p}_+} \hat{\phi}_{\mathbf{p}}^* e^{-i(\omega_{\mathbf{p}_+} - \omega_{\mathbf{p}})t} - (p_-^2 - p^2) \hat{\phi}_{\mathbf{p}_-}^* \hat{\phi}_{\mathbf{p}} e^{-i(\omega_{\mathbf{p}} - \omega_{\mathbf{p}_-})t} \right) \tag{1.36}$$

$$\partial_t \hat{\phi}_{\mathbf{p}_+} = \frac{p_y q}{1+p_+^2} (p^2 - q^2) \hat{\phi}_{\mathbf{p}} \hat{\phi}_{\mathbf{q}} e^{-i(\omega_{\mathbf{p}} - \omega_{\mathbf{p}_+})t} \tag{1.37}$$

$$\partial_t \hat{\phi}_{\mathbf{p}_-} = -\frac{p_y q}{1+p_-^2} (p^2 - q^2) \hat{\phi}_{\mathbf{p}} \hat{\phi}_{\mathbf{q}}^* e^{-i(\omega_{\mathbf{p}} - \omega_{\mathbf{p}_-})t}. \tag{1.38}$$

We now look for harmonic solutions to this equation of the form

$$\hat{\phi}_{\mathbf{q}} = A_{\mathbf{q}} e^{-i\Omega_{\mathbf{q}} t} \tag{1.39}$$

$$\hat{\phi}_{\mathbf{p}_\pm} = A_\pm e^{-i\Omega_\pm t}, \tag{1.40}$$

which requires $\Omega_+ = \omega_{\mathbf{p}_+} - \omega_{\mathbf{p}} - \Omega_{\mathbf{q}}$ and $\Omega_- = -\omega_{\mathbf{p}_-} + \omega_{\mathbf{p}} - \Omega_{\mathbf{q}}^*$. There are nontrivial solutions of such a form when

$$\Omega_{\mathbf{q}}(\omega_{\mathbf{p}} - \omega_{\mathbf{p}_+} + \Omega_{\mathbf{q}})(\omega_{\mathbf{p}} - \omega_{\mathbf{p}_-} - \Omega_{\mathbf{q}}) = \tag{1.41}$$

$$p_y^2 q^2 |\hat{\phi}_{\mathbf{p}}|^2 \frac{p^2 - q^2}{1 + q^2} \left[\frac{p_-^2 - p^2}{1 + p_-^2}(\omega_{\mathbf{p}} - \omega_{\mathbf{p}_+} + \Omega_{\mathbf{q}}) - \frac{p_+^2 - p^2}{1 + p_+^2}(\omega_{\mathbf{p}} - \omega_{\mathbf{p}_-} - \Omega_{\mathbf{q}}). \right] \tag{1.42}$$

This is a cubic polynomial in $\Omega_{\mathbf{q}}$; to leading order in q, its discriminant is negative when $1 - 3p_x^2 + p_y^2 > 0$ and

$$|\hat{\phi}_{\mathbf{p}}|^2 > \frac{\kappa^2(1 - 3p_x^2 + p_y^2)}{2p^2(1 + p^2)^4}. \tag{1.43}$$

Hence, if the primary drift wave amplitude is sufficiently large and $q < 1$, then $\Omega_{\mathbf{q}}$ will have an imaginary part, and there will be an exponentially growing solution for the zonal mode amplitude.

This calculation demonstrates that the zonal flow instability is a *modulational instability*: in response to a small, inhomogeneous seed zonal flow, nonlinear drift wave interactions reinforce the inhomogeneity.

Another, equivalent perspective on zonal flow formation is that the turbulent Reynolds stress takes the form of a *negative viscosity*, so that

$$\partial_t V_y \simeq -\chi_T \partial_x^2 V_y + \text{other terms.} \tag{1.44}$$

Here, χ_T is positive-definite and scales with the turbulence intensity. A flow evolving according to such an equation will be unstable to modulations. This negative viscosity picture will be discussed in greater detail in Ch. 2.

Turbulence-generated zonal flows tend to self-organize into a quasiperiodic structure with a characteristic, mesoscopic lengthscale. The physics which determines this lengthscale is complex and largely an open problem.

1.3.2 Turbulence spreading and avalanching

We have so far discussed drift-wave turbulence as being driven by a linear instability mechanism, that is, turbulence results when the density (or temperature) gradient exceeds a critical threshold. In the tokamak, experiments indicate the presence of turbulent fluctuations *even in linearly stable regions* where the gradient is below threshold. Simulations and theoretical arguments suggest that this is due to *turbulence spreading*, a process whereby turbulence self-propagates through nonlinear interactions. In particular, turbulence originating in unstable regions can invade nearby stable regions. Through spreading, the local gradient alone is insufficient to determine the turbulence saturation level. The turbulence level, in turn, determines the level of anomalous transport, so spreading results in a *spatially nonlocal* relationship between transport and the associated profile gradient.

Spreading, like zonal flow generation, can be viewed as a multiscale interaction: now, the large scale $\mathbf{q}$ is an *envelope* scale for the drift wavetrain. Applying the methods of statistical closure to the convective nonlinearity will generally yield an equation of the form [DII10]

$$\partial_t |\phi_{\mathbf{k}}|^2 = \sum_{\mathbf{k}'+\mathbf{k}''=\mathbf{k}} \left[\hat{z}\cdot(\mathbf{k}\times\mathbf{k}')\right]^2 R(\mathbf{k},\mathbf{k}',\mathbf{k}'') \tag{1.45}$$

$$\times \left[M^2_{\mathbf{k},\mathbf{k}',\mathbf{k}''}|\phi_{\mathbf{k}'}|^2|\phi_{\mathbf{k}''}|^2 - 2\Gamma_{\mathbf{k},\mathbf{k}',\mathbf{k}''}|\phi_{\mathbf{k}}|^2|\phi_{\mathbf{k}'}|^2\right], \tag{1.46}$$

where M and Γ are coefficients and $R(\mathbf{k},\mathbf{k}',\mathbf{k}'')$ is the resonance function which determines the triad interaction time — for weak turbulence, this is simply $\delta(\omega_{\mathbf{k}} - \omega_{\mathbf{k}'} - \omega_{\mathbf{k}''})$. The second term in the summand, identified as the nonlinear damping, gives rise to spreading effects when a radial envelope is introduced and treated on an equal footing with the mode scale, i.e. $\mathbf{k} \to \mathbf{k} - i\partial_x\hat{x}$. Among other terms, one obtains

$$\partial_t |\phi_{\mathbf{k}}|^2 = \partial_x(D_x\partial_x|\phi_{\mathbf{k}}|^2) + \ldots, \tag{1.47}$$

where

$$D_x = 2 \sum_{\mathbf{k}'+\mathbf{k}''=\mathbf{k}} k_y^2 \Gamma_{\mathbf{k},\mathbf{k}',\mathbf{k}''} R(\mathbf{k},\mathbf{k}',\mathbf{k}'') |\phi_{\mathbf{k}'}|^2. \tag{1.48}$$

This has the form of nonlinear diffusion of turbulence intensity — the diffusivity itself scales with the turbulence intensity, reflecting the fact the turbulence is self-propagating. Thus, spreading is often represented in simple models by a nonlinear diffusion term. One popular such model takes the form of a "Fisher equation"

$$\partial_t I = \gamma_0 I - \gamma_{nl} I^2 + \partial_x (D(I) \partial_x I), \tag{1.49}$$

where I represents the local turbulence intensity, γ_0 is the local linear growth (or damping), and γ_{nl} models nonlinear damping. The diffusivity is often taken to be $D(I) = D_0 I^\beta$ for some $0 < \beta \leq 1$. This model and extensions thereof are the subject of Ch. 3.

Finally, we mention *avalanching*, a phenomenon which is closely related to turbulence spreading. Avalanches, which can carry a large fraction of the total transport in a magnetic confinement device, are fast transport events which result from the sequential firing of localized cells or modes, akin to a "domino effect." They are associated with the rapid relaxation of an unstable configuration into a metastable configuration and thus play a role in the theory of self-organized criticality [BTW87].

Avalanching can, again, be viewed as the result of multiscale coupling: now, a small mode scale couples with a mesoscale avalanche scale which measures the spatial extent of the transport event, i.e. the range over which local cells are excited. Such a transport event will inevitably lead to turbulence spreading, so the concepts are inextricably linked.

1.4 MHD, on a β-plane and otherwise

Zonal flows appear in a a variety of plasma systems. Another such system of interest to this dissertation is β-plane magnetohydrodynamics (MHD), which is used as a simple model for certain solar systems.

Let us first briefly review (ideal, incompressible) MHD more broadly. MHD is a set of equations which model a plasma as a single, quasineutral fluid (rather than a dual fluid of electrons and ions). Thus, electric fields are the result of magnetic induction (rather than charge imbalance) through Faraday's law

$$\nabla \times \mathbf{E} = -\frac{1}{c}\partial_t \mathbf{B}. \tag{1.50}$$

Combining this with Ohm's law

$$\eta \mathbf{J} = \mathbf{E} + \frac{1}{c}\mathbf{v} \times \mathbf{B}, \tag{1.51}$$

the nonrelativistic Ampère's law

$$\nabla \times \mathbf{B} = \frac{4\pi}{c}\mathbf{J}, \tag{1.52}$$

and Gauss' law

$$\nabla \cdot \mathbf{B} = 0 \tag{1.53}$$

yields the *magnetic induction equation*

$$\partial_t \mathbf{B} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \frac{\eta c^2}{4\pi}\nabla^2 \mathbf{B}. \tag{1.54}$$

For an incompressible flow with $\nabla \cdot \mathbf{v} = 0$ and uniform density ρ, this can be rewritten

$$\partial_t \mathbf{B} + \mathbf{v} \cdot \nabla \mathbf{B} = \mathbf{B} \cdot \nabla \mathbf{v} + \frac{\eta c^2}{4\pi}\nabla^2 \mathbf{B}. \tag{1.55}$$

The magnetic induction equation, together with the incompressibility condition, Gauss' law, and the momentum equation

$$\partial_t \mathbf{v} + \mathbf{v} \cdot \nabla \mathbf{v} = -\frac{\nabla p^*}{\rho} + \frac{1}{4\pi\rho} \mathbf{B} \cdot \nabla \mathbf{B} + \nu \nabla^2 \mathbf{v} \tag{1.56}$$

(where $p^* = p + B^2/8\pi$ is the total pressure), form the basic equations of ideal, incompressible MHD. There are many ways to relax the assumptions made above, each yielding a more complicated version of this system, but ideal, incompressible MHD already gives rise to an interesting array of phenomena.

One of the most fundamental principles of ideal MHD is that, neglecting resistivity, the magnetic field lines are *frozen in* to the fluid flow and must move along with it, even as they determine the flow through the Lorentz force. This is known as *Alfvén's theorem*. Ideal MHD also conserves the total energy

$$E = \int d^3\mathbf{x} (v^2 + \frac{B^2}{4\pi}) \tag{1.57}$$

as well as two topological invariants called the magnetic helicity

$$H_M = \int d^3\mathbf{x} \mathbf{A} \cdot \mathbf{B} \tag{1.58}$$

(where A is the vector potential) and the cross helicity

$$H_C = \int d^3\mathbf{x} \mathbf{v} \cdot \mathbf{B}. \tag{1.59}$$

In the presence of a mean magnetic field $\mathbf{B}_0$, ideal, incompressible MHD features linear excitations called *Alfvén waves*, which have dispersion relation $\omega = \pm \frac{\mathbf{B}_0 \cdot \mathbf{k}}{\sqrt{4\pi\rho}}$. Alfvén waves are transverse waves which propagate parallel to the magnetic field, with group velocity $v_A = B_0/\sqrt{4\pi\rho}$ (called the Alfvén speed).

One fascinating application of MHD to astrophysics, a topic of ongoing research, is the *dynamo problem*: how can a turbulent plasma flow self-consistently generate and sustain a large-scale, long-lived magnetic field, such as the Sun's or the Earth's? Answering this requires understanding the large-scale turbulent emf $\langle \tilde{\mathbf{v}} \times \tilde{\mathbf{B}} \rangle$ generated by the interactions of small-scale fluctuations—yet another example of turbulent transport through multiscale interaction.

In two dimensions, we can introduce a streamfunction ψ and a scalar potential A, with $\mathbf{v} = \hat{z} \times \nabla\psi$ and $\mathbf{B} = \hat{z} \times \nabla A$. Then the pressure term can be eliminated from the momentum equation by taking the curl, and the induction equation can be simplified by "uncurling." The incompressible 2D MHD equations are then

$$\partial_t \nabla^2 \psi + (\hat{z} \times \nabla\psi) \cdot \nabla\nabla^2\psi = (\hat{z} \times \nabla A) \cdot \nabla\nabla^2 A + \nu\nabla^4\psi \tag{1.60}$$

$$\partial_t A + (\hat{z} \times \nabla\psi) \cdot \nabla A = \eta\nabla^2 A. \tag{1.61}$$

This system is also relevant in three dimensions when there is an exceptionally strong mean magnetic field along $\hat{z}$.

The β-plane system is obtained by adding a term $\beta\partial_x\psi$ to the LHS of the vorticity equation. Physically, this amounts to imagining the system is a patch, at latitude λ, on the surface of a planet (or star) rotating at frequency Ω (x now represents meridian and y represents latitude). The β term then represents a linear approximation of the local variation of the Coriolis force with latitude, and $\beta = 2\Omega\cos\lambda/R$, where R is the radius of the planet. β is called the *Rossby parameter*.

In the absence of a magnetic field, the vorticity equation on a β-plane is equivalent to the HM equation in the limit that $\rho_s \to \infty$, so it is unsurprising that when this equation is forced, spontaneous zonal flows will result. The magnetic field complicates matters by opposing the zonal flow generation: the (collisionless) equation for the mean flow is now

$$\partial_t V_x = \partial_x(\langle \partial_x\tilde{\psi}\partial_y\tilde{\psi} \rangle - \langle \partial_x\tilde{A}\partial_y\tilde{A} \rangle), \tag{1.62}$$

and a *Maxwell stress* now counteracts the Reynolds stress. In the presence of a sufficiently strong mean magnetic field in the plane, the turbulence becomes “Alfvénized,” and the Maxwell stress conspires to almost precisely cancel the Reynolds stress. No zonal flows form. In general, there is a transition between this Alfvénized regime and a Rossby-wave regime where zonal flows are able to form, depending on the relative strength of the magnetic field and the Rossby parameter.

1.5 Overview of remaining chapters

This dissertation focuses on three topics in plasma turbulence, loosely connected through the theme of multiscale interaction and transport. These are accordingly divided into three chapters.

In Ch. 2, we focus on the turbulent dynamics of the HW system, in two dimensions and in the regime where zonal flows are abundant (this requires an adiabaticity parameter $\alpha > 1$ and a gradient drive that is supercritical but not too strong). In this regime, the system is quasi-one-dimensional due to poloidal symmetry. Thus, we seek a mean-field description for the dynamics of the zonally averaged density, flow, and turbulence intensity. To accomplish this, we use *deep learning* applied to an ensemble of direct numerical simulations (DNS) of the system. In particular, deep learning is used to infer from the simulations approximate models for the turbulent Reynolds stress and particle flux, as functions of the zonally averaged variables and their radial derivatives. Symmetry and other physical constraints are used to help inform the structure of the deep learning model. We compare the deep-learning-based model to simple analytic models and, using numerical solution of the deep-learning-based model, to the original DNS. The deep learning approach is devised so that it may be generalized and applied to study other systems which may be more complicated than HW.

In Ch. 3, we turn our attention to turbulence spreading. We discuss weaknesses of the Fisher model: for instance, it predicts only a feeble degree of penetration into stable regions. These weaknesses motivate the presentation of a new model which is based on bistable/subcritically unstable turbulence, which we support by pointing to a number of studies which suggest that plasma turbulence can nonlinearly self-sustain even when linearly damped. This model is thoroughly analyzed and connected to other models, including models for avalanches and for the transition to turbulence in pipe flow. The key results of our analysis are that (a) the bistable model supports far more robust penetration into stable regions and is thus arguably more suitable than the Fisher model and (b) there is a threshold in spatial extent and amplitude for a puff of turbulence to grow. We liken this latter behavior to a rudimentary model of an avalanche, and we derive an approximation for the threshold using a simple physical argument.

In Ch. 4, we study the β-plane MHD model with a mean magnetic field in the plane. We note that cross-helicity conservation is broken by the β-term, and derive an estimate for the stationary mean cross-helicity using conservation of mean squared magnetic potential. Then, we use weak turbulence theory (suitably generalized to a system with multiple linear modes) to show that, in a stationary state, the time-averaged cross-helicity spectrum determines momentum transport. This result is also a precise mathematical version of the statement that the turbulence is magnetic at scales smaller than a critical lengthscale (called the magnetic Rhines scale), and kinetic at larger scales. We supplement our results with computer simulations. We find that the cross-helicity is unsurprisingly largest when the system is close to transition (i.e. between directly cascading Alfvénic turbulence and inversely cascading Rossby-wave turbulence), which occurs around $\beta \sim b_0 k_f^2$, where k_f is the forcing scale.

Finally, in Ch. 5, we conclude with a summary of the results of the preceding chapters and a critical, holistic discussion of the lessons learned.

Chapter 2

Learning how structures form in drift-wave turbulence

2.1 Introduction

Drift-wave turbulence [Hor99] is a ubiquitous phenomenon in magnetic fusion devices which plays a central role in anomalous transport. Endemic to drift-wave turbulence is structure formation via nonlinear feedback loops. The most familiar such structure is the zonal flow (ZF)[DIIH05a], an axisymmetric flow with $\omega \simeq 0$. The zonal flow is a secondary structure, driven by turbulence, which itself suppresses turbulent transport and whose formation is responsible for the confinement-improving L-H transition [KD03]. In addition, features such as quasiperiodic staircases [DPDG^{+}10, DPHG^{+}15, ANB^{+}19] are known to form in profiles and themselves impact transport and confinement.

The formation of nonlinear structures in drift-wave turbulence is the result of feedback loops resulting from the interaction of at least three major players: the profile, the ZF, and the turbulence intensity field. The profile drives the turbulence via (primary) instability; the (secondary) ZF is driven by the turbulence and in turn suppresses it via eddy shearing; and the

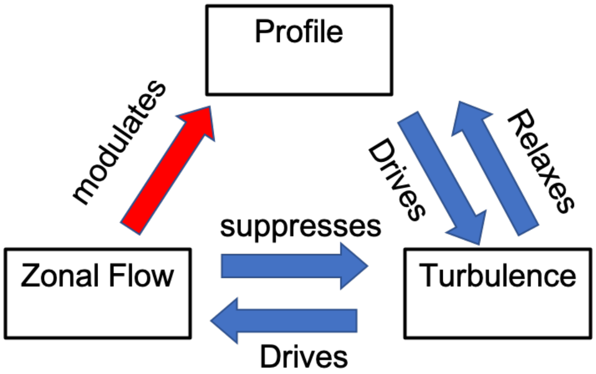


Figure 2.1: Cartoon depicting the basic feedback loops in the drift-wave/ZF system. The interaction between ZF and profile is especially pertinent to this work and is highlighted in red.

turbulence induces a flux which relaxes the driving profile, tending to quench the instability (see Fig. 2.1). Moreover, we will see that the presence of ZFs tends to induce modulations in the profile.

Understanding these feedback loops (and all other important aspects of the transport and nonlinear dynamics) demands the study of turbulent fluxes produced by cross-correlations between the fluctuations. For example, it is the production of a Reynolds stress $\Pi = \langle \tilde{v}_r \tilde{v}_\theta \rangle$ that gives rise to the ZF. (Here and throughout, the angle brackets refer to an average over directions of symmetry and the tilde indicates the local deviation from that average, e.g. $\tilde{\phi} = \phi - \langle \phi \rangle$.) However, the calculation of turbulent fluxes is a challenging problem analytically, whose solution from first principles always requires the use of successive — and sometimes questionable — approximations, such as the introduction of a small parameter, or a closure for higher-order moments.

A classical approach to modeling turbulent fluxes is based on the local mixing-length theory (following Prandtl's work on turbulent jets [Pra25]). In such a prescription, the turbulent transport is characterized by an effective diffusivity ℓ_{mix}^2/τ, where the mixing length ℓ_{mix} is the correlation length associated with turbulent convection, and τ is a characteristic timescale. The canonical mixing-length model is Kadomtsev's estimate for the particle flux [Kad65]

$$\Gamma \simeq -\frac{\gamma_{\mathbf{k}}}{k_\perp^2}\frac{d\langle n\rangle}{dx}, \tag{2.1}$$

where $\gamma_{\mathbf{k}}$ is the linear growth rate.

More generally, one can seek a local mean-field theory (MFT) $\mathscr{M}$ that can predict the instantaneous, zonally-averaged flux at a given radius and time as a function of other instantaneous, zonally-averaged variables associated with the profiles, flow, and turbulence. The applicability of MFT is grounded in the (approximate) poloidal symmetry of the problem. Formally, one choses a collection of n spatiotemporally-varying fields $\psi_i(\mathbf{x},t)$ and seeks a map

$$\mathscr{M}_\xi : (\langle\psi_1\rangle,\ldots,\langle\psi_n\rangle)|_{r_0,t_0} \mapsto \langle\tilde{v}_r(r_0,t_0)\tilde{\xi}(r_0,t_0)\rangle \tag{2.2}$$

outputting the turbulent flux of ξ at a radius and time (r_0,t_0). (Note that in certain systems, one must consider additional contributions to the flux beyond this convective part.) Examples of ψ_i in a real system might be the electron and ion temperatures, the electron and ion densities, components of the electric field, the poloidal flow shear, and radial derivatives of these fields. Choosing the ψ_i requires input of physics knowledge or intuition, such as symmetries, and truncation of the chain of derivatives at some order.

To give a concrete example, we may consider the Hasegawa-Wakatani (HW) system for the potential ϕ and electron density n [HW83, WH84]

$$\partial_t n + \{\phi, n\} = C(n-\phi) \tag{2.3}$$

$$\partial_t \nabla_\perp^2\phi + \{\phi, \nabla_\perp^2\phi\} = C(n-\phi). \tag{2.4}$$

Here, $\{\cdot,\cdot\}$ is the Poisson bracket, $C = \frac{T_e}{m_e \nu_{ei}\rho_s c_s}\partial_z^2$ is the adiabatic operator, and we have used the usual normalizations $\ln(n/n_0) \to n, \phi \to e\phi/T_e, x \to \rho_s x, t \to t/\omega_{ci}$. [To be clear, T_e (T_i) is the electron (ion) temperature, m_e (m_i) is the electron (ion) mass, ν_{ei} is the parallel electron-ion collision frequency, ω_{ci} is the ion gyrofrequency, $\rho_s = \sqrt{\frac{T_e}{m_i}}\omega_{ci}^{-1}$, and $c_s^2 = T_e/m_i$.] Dissipation terms have been neglected. This system conserves two independent quadratic invariants: the energy $E = \int d^3\mathbf{x}\,(n^2 + (\nabla\phi)^2)$ and the potential enstrophy (PE) $W = \int d^3\mathbf{x}\,(n - \nabla^2\phi)^2$.

The HW system, which models resistive drift-wave turbulence, is the simplest *realistic* paradigm to study, in that it features a linear instability mechanism and profile evolution, in contrast to the simpler Charney-Hasegawa-Mima equation [Cha71, HM77] (the adiabatic limit of HW). It is useful to separate the mean and fluctuating parts of Eqs. (2.3–2.4) to obtain

$$\partial_t\tilde{n} + N'\partial_y\tilde{\phi} + V_y\partial_y\tilde{n} = C(\tilde{n} - \tilde{\phi}) \tag{2.5}$$

$$\partial_t\tilde{\nabla}_\perp^2\phi - V_y''\partial_y\tilde{\phi} + V_y\partial_y\nabla_\perp^2\tilde{\phi} = C(\tilde{n} - \tilde{\phi}) \tag{2.6}$$

$$\partial_t N + \partial_x\Gamma = 0 \tag{2.7}$$

$$\partial_t V_y' - \partial_x^2\Pi = 0, \tag{2.8}$$

where we have used the Taylor identity [TS15] in obtaining the last equation (see App. A). Here, $N = \langle n\rangle, V_y = -\partial_x\langle\phi\rangle$. A prime indicates an x derivative. Note that we have also approximated the nonlinearities by their mean values; that is, we have set $\widetilde{\tilde{n}\partial_y\tilde{\phi}} = \widetilde{\partial_x\tilde{\phi}\partial_y\tilde{\phi}} = 0$.

At this stage, a model for the turbulent fluxes is needed. One possibility is to impose a mixing-length ansatz; for example, Ashourvan and Diamond [AD17, AD16] proposed a model where the fluxes are proportional to mean gradients and the turbulence intensity, e.g. $\Gamma = -c\ell_{mix}^2\varepsilon N'$, where $\varepsilon = \langle(\tilde{n} - \nabla_\perp\tilde{\phi})^2\rangle$ is the turbulent potential enstrophy (PE), with an ansatz for the mixing length ℓ_{mix} based on turbulence bistability. (This model is discussed in detail in Sec. 2.4.5.) While it successfully generates ZFs, staircases, and other features, their model (along with all other mixing-length models) is heuristic and cannot be derived from first principles. In this work, we suggest and explore an alternative *data-driven* approach for mean-field modeling which uses *deep learning* to infer dependencies of fluxes on mean quantities of interest. This allows us to obtain a reduced model directly from the exact dynamical equations while circumventing the need for challenging analytical calculations.

Deep learning [LBH15] refers to the use of algorithms which process data through multiple layers in order to learn abstract representations of the data. Such algorithms exist in many forms, but in this work we will use one of the simplest, a feedforward deep neural network (DNN), also called a multi-layer perceptron (MLP). The utility of DNNs to our work lies in their ability to approximate arbitrary continuous multivariate functions, as stated by the numerous variations of the *universal approximation theorem* [Hor91, LLPS93, LPW$^+$17], as well as their resilience to vast amounts of noise in the dependent variable — in certain applications, DNNs have been shown to train successfully even when as much as 99% of the data are randomly labeled [RVBS17].

The scheme of the deep learning approach is as follows. The exact turbulent model equations are first solved numerically, over a broad range of initial conditions. The data thereby generated represent a map of the form (2.2), albeit a highly noisy one, due to both intrinsic turbulent noise and deviations from the mean-field model. Finally, supervised learning[1] is used to filter that noise and distill an arbitrary, deterministic model for the fluxes, free of any imposed functional form. We impose only a minimal set of assumptions — the existence of a local mean-field model for the fluxes, which obey the symmetries guaranteed by the underlying equations, along with a choice of parameters. Our approach, a form of fully nonlinear, nonparametric regression, finds that model which best explains the mean-field dynamics, taking mean field theory "to the end of the road." It may serve to verify an existing model or to probe a poorly-understood system and uncover the important emergent nonlinear dynamics.

In this work, we apply this idea to the 2-D HW system as a test of concept. 2-D HW is a natural testing ground as it is reasonably analytically tractable and can be solved fast enough numerically to quickly generate training data. We extract models for both the particle flux Γ and the Reynolds stress Π. The deep learning method highlights the feedback of the ZF on the driving

---

[1] "Supervised learning" refers to machine learning whose training data consist of complete input-output pairs — that is to say, we know from our simulations the correct flux corresponding to each set of inputs.

profile via an “off-diagonal” particle flux proportional to the gradient of mean vorticity or shear. In particular, it finds that this rarely-discussed effect is *significant* — in this system, moreso than the direct effect of the shear itself. We support this finding by a simple quasilinear calculation with mean flow. We will see that the off-diagonal flux straightforwardly leads to staircase formation, in a manner that is distinct from previous models based on bistability associated with shearing or a Rhines scale.

Meanwhile, the DNN learns a model for the Reynolds stress consisting of negative diffusion stabilized by a nonlinearity and a hyperdiffusion. This result agrees well with a simple calculation from the wave-kinetic equation in the presence of a background flow, in addition to previous theoretical work.

These basic results appeared previously in Ref. [HD20]. In this work, we significantly expand on that paper, discussing in detail the feature formation processes in the HW system, presenting additional findings from the deep learning model, directly comparing the learned particle flux to one obtained using an ansatz spectrum, obtaining analytically the nonlinear dependence of the Reynolds stress on the mean vorticity, and numerically simulating a reduced model based on the findings of the deep learning method.

The paper is organized as follows: in Sec. II, we give details on our numerical solutions and the deep learning model. In particular, we emphasize the importance of imposing symmetry constraints on the deep learning model. In Sec. III, results for the learned particle flux and Reynolds stress are presented and discussed. Finally, in Sec. IV, these results are compared to theoretical calculations. We also introduce a 1-D reduced model for the interaction of the mean field density, flow, and turbulence intensity which is based on our findings. This model is solved numerically and compared to the mixing length model of Ashourvan and Diamond.

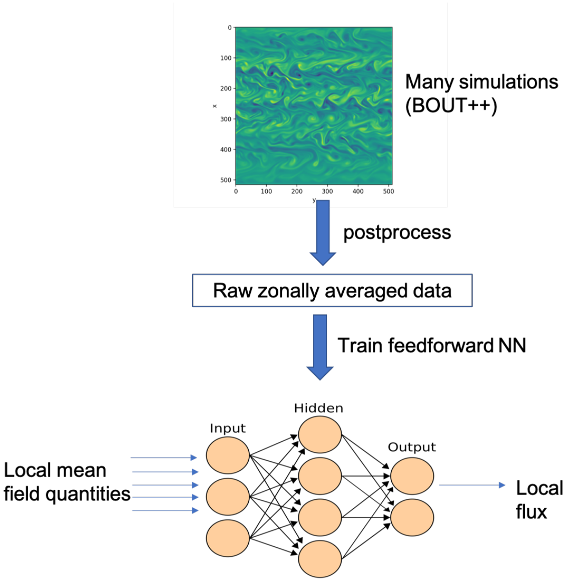


Figure 2.2: Basic schematic of the deep learning method. Simulations are ran, post-processed, and then fed into a feedforward neural network.

2.2 Methods

2.2.1 Numerical solution of 2-D HW

The workflow of the method is illustrated in Fig. 2.2. We first perform direct numerical simulation (DNS) of the (modified) HW system [NBD07] on a 2-D slab

$$\partial_t n + N'(x)\partial_y \phi + \{\phi, n\} = \alpha(\tilde{\phi} - \tilde{n}) - D\nabla^4 n \tag{2.9}$$

$$\partial_t \nabla^2 \phi + \{\phi, \nabla^2 \phi\} = \alpha(\tilde{\phi} - \tilde{n}) - \mu\nabla^2\phi + D\nabla^6\phi. \tag{2.10}$$

The adiabatic operator has been replaced with a constant which we fix at $\alpha = 2$ in all simulations. This places us in the weakly adiabatic regime; in future work, we will relax this restriction. The tildes on the RHS are important for ZF generation and respect the fact that zonal components do not contribute to the parallel current [DH93]. The background gradient drive is varied from run to run but is generally chosen as either constant ($N'(x) = \kappa$) or linearly varying ($N'(x) = \beta x$). The gradient drive is chosen to be large enough to exceed the Dimits shift regime [DBB$^+$00] —

for small, (linearly) supercritical gradient drive, undamped zonal flows dominate and the system is nonturbulent, leading to a nonlinear upshift in the instability threshold. The hyperdiffusivity is fixed at $D = 10^{-4}$ and the linear flow damping at $\mu = 10^{-2}$, and the box size is such that our effective ρ_* is $1/51.5$. Note that dissipation terms are *small* (compared to unity, or upon redimensionalizing $\mu \ll \omega_{ci}$ and $D \ll \rho_s^4 \omega_{ci}$), and included primarily for stability reasons.

Using the BOUT++ framework [DUX$^+$09], this system is solved on a square 512×512 spatial grid using the Karniadakis time-stepping algorithm [KIO91]. We use periodic boundary conditions in y for all variables. In the x direction, we employ homogeneous Neumann boundary conditions for n and homogeneous Dirichlet boundary conditions for ϕ and $\nabla^2\phi$. A small broad-spectrum fluctuation is initialized in the vorticity to start up the instability, and in some simulations a background ZF is initialized. The data are outputted to file at time intervals of size $\tau = 1$.

A total of 32 runs with different initial conditions are used in the training data, each with 2000 outputted timesteps (though the first ten are discarded). In detail, ten simulations have a uniform background gradient $0.75 \leq N' \leq 3$, seven have a linearly varying gradient with $1 \leq \beta \leq 5$, and the remaining fifteen have both a uniform gradient $1 \leq N' \leq 1$ and a background flow $V_y = v_0 \cos(2\pi n x / L_x)$ with $n = 1, 2, 3$.

2.2.2 Post-processing

Next, zonally averaged quantities of interest—namely, the turbulence intensity (here represented by the turbulent PE ε—see Sec. 2.2.4), the density gradient N', vorticity U and its derivatives U' and U'', and the fluxes Γ and Π—are computed from the aggregated numerical solution data. The radial dimension is coarse-grained: the data are also averaged over a small window of four radial grid points, with derivatives computed using finite differences. Thus, each simulation run produces, for each flux, $128 N_t = 254720$ data points representing the map (2.2), where N_t is the number of outputted timesteps, and a data point consists of a tuple of inputs $(\varepsilon, N', U, U', U'')$ equipped with a corresponding turbulent flux. Note that our underlying assumption of space-time locality means each simulation generates a wealth of training data.

2.2.3 DNN training

Finally, the data are used to train a simple feedforward DNN. To ease the burden of the jargon in this section, let us first review the notion of a DNN and the training procedure (for more information, see for example Ref. [MBW$^+$19], an introduction to machine learning intended for physicists). A DNN represents a generic map which transforms an input vector $\mathbf{x}^0$ by a sequence of nonlinear maps called "hidden layers." For our purposes, we consider an MLP, where each hidden layer transforms the output of the previous layer as

$$x_j^{i+1} = \sigma\left(\mathbf{w}_j^i \cdot \mathbf{x}^i + b_j^i\right). \tag{2.11}$$

Here, $\mathbf{w}_j^i$ and $\mathbf{b}^i$ are trainable parameters referred to, respectively, as *weights* and *biases*, and σ is a specified nonlinear map called the activation function. Common choices for the activation function are the hyperbolic tangent, the sigmoid $1/(1+e^{-x})$, and the rectified linear unit (ReLU) $\max(x,0)$. The upper index labels the layer, and the lower index labels the *neuron* or *unit*; the activation function thus, in a sense, specifies the response of each neuron to the input. An output layer with no activation function transforms the output of the final hidden layer to yield a single number which represents a flux; this is the output of the DNN. The weights and biases are trained to minimize a loss function $L(\mathbf{w},\mathbf{b},\{\mathbf{x}^0\},\{y^*\})$, where $\{\mathbf{x}^0\}$ is the set of training inputs; the loss quantifies the deviation of the DNN prediction from the corresponding fluxes which were actually seen in simulation $\{y^*\}$.

"Training" refers to this process of optimizing the weights and biases using the simulation data. The most common approach is to use some version of *stochastic gradient descent* (SGD), which differs from standard gradient descent by estimating the gradient of the loss function using a small batch of training points rather than the entire set. During each batch, the weights and biases are incremented according to this estimate, e.g. $\mathbf{w}_j^i \to \mathbf{w}_j^i - \eta \nabla_{\mathbf{w}_j^i} L$, for some learning rate η. A full set of batches constituting the full dataset is referred to as an *epoch*. Training a neural network usually requires many epochs. Most commonly, data is split into "training" and "validation" sets; the training set is used for SGD, and training is terminated when the loss, as measured on the validation set, ceases to improve for some specified number of epochs. This is called *early stopping*. The partitioning into training and validation sets helps prevent *overfitting*, wherein the DNN too precisely reproduces the training data without properly generalizing to unseen data. Overfitting is akin to using a high-degree polynomial which passes through all points as a fit to data that could have been well-modeled by a line.

Building a DNN model requires the choice of several *hyperparameters* which specify its structure and training procedure, such as the learning rate η, the number of hidden layers, the number(s) of neurons in each hidden layer, the activation function, the size of the training batches, and details of the loss function.

Our DNN uses three hidden layers with eight neurons each. We employ the "exponential linear unit" [CUH15]

$$f(x) = \begin{cases} x, & x \geq 0 \\ e^x - 1, & x < 0, \end{cases} \tag{2.12}$$

a smoother alternative to the ReLU, as our hidden layer activation function. Batch normalization (BN) [IS15], which ensures the distributions of data that are inputted to the hidden layers have unit variance and zero mean, is applied after each hidden layer. BN is widely used to help accelerate and stabilize training. We trained on the aggregate simulation data, randomly separated into training and validation sets, in batches of size 256 using the Adam algorithm for SGD [KB14]. To ensure that the result does not depend on the choice of training and validation sets, this data partitioning is performed ten times, resulting in ten independently trained models, and their outputs are averaged. In separate training runs, we partitioned the data further, excluding from the training and validation sets a "test set" corresponding to a specific range of initial N', which constituted about 15% of the data. We checked that the model still properly trained and performed well on the excluded data.

The optimization was performed with respect to the loss function

$$L = \sum_i \ln\left(\cosh(y_i^* - f_W(\mathbf{x}_i))\right) + \lambda ||W||^2, \tag{2.13}$$

where W is the matrix of network weights, $\mathbf{x}_i$ is the set of inputs (U, N', etc.) for the i-th data point, y_i^* is the corresponding flux, f_W is the map encoded by the DNN which predicts the flux, $||\cdot||$ is the Frobenius norm, and $\lambda = 10^{-5}$. We choose this "logcosh" loss in an effort to suppress the effect of noise, as it is *quadratic* (and smooth) in the error for small arguments, but asymptotically *linear* for large arguments. This way, large outliers are not penalized too heavily. The (standard) L^2 regularization term $\propto ||W||^2$ is aimed at reducing model complexity. This, as well as early stopping after two training epochs without an improvement in the validation accuracy, is used to deter overfitting.

It is natural to ask how much freedom we have to choose the hyperparameters; the efficacy of a nonparametric method like deep learning should not depend too heavily on the precise structure of the neural network. Indeed, we find that the result is rather robust to variations in the number of neurons and hidden layers, as long as the network has sufficient complexity/representation power. On the other hand, certain parameters required "tuning" for performance. For instance, setting λ too large overpenalizes complexity and yields unphysical results.

Above all, we emphasize that the DNN training is simply a sophisticated form of nonparametric regression which minimizes L, a representation of the error in predicting the flux.

2.2.4 Feature selection and symmetry constraints

In order to successfully train the model, some physics input is required. At a minimum, one must face the problem of feature selection, i.e. choose the inputs on the LHS of Eq. (2.2). This demands some understanding of which mean field quantities the turbulent fluxes are likely to directly depend on. Exact symmetries of the 2-D HW system are useful here: it is invariant under constant shifts $n \to n + n_0$ and $\phi \to \phi + \phi_0$ as well as Galilean poloidal boosts of the form

$$\begin{cases} \phi \to \phi + v_0 x \\ y \to y - v_0 t. \end{cases} \tag{2.14}$$

These symmetries preclude direct dependence of the flux on $\langle n \rangle$, $\langle \phi \rangle$, or $\partial_x \langle \phi \rangle$. We choose N', U, U', U'' as independent variables (U'' dependence is included in anticipation of a stabilizing hyperdiffusion term in the Reynolds stress). Higher-order derivatives could be included in principle, but this introduces numerical noise due to the finite differencing. Moreover, it is preferable to minimize the number of input parameters, as increasing the dimensionality rapidly makes training more difficult, while complicating and obfuscating the model's dependencies.

A proxy for the local turbulence intensity is also needed as an independent variable; we choose the turbulent PE $\varepsilon = \langle (\tilde{n} - \nabla_\perp \tilde{\phi})^2 \rangle$ [AD16, AD17]. We stress that other choices, such as the potential fluctuation intensity $\langle \tilde{\phi}^2 \rangle$ and the turbulent energy $E = \langle \tilde{n}^2 + (\nabla \tilde{\phi})^2 \rangle$, are equally valid. However, in the adiabatic regime $\alpha > 1$ which we consider, we have $\varepsilon_{\mathbf{k}} \simeq (1 + k^2)^2 |\phi_{\mathbf{k}}|^2 \simeq (1 + k^2) E_{\mathbf{k}}$, so that in the case of a sharply peaked spectrum, the local energy, fluctuation intensity, and PE only differ by constant factors.

The 2-D HW system also obeys a group of reflection symmetries (isomorphic to the Klein four-group), whose nontrivial elements are

$$x \to -x, y \to -y; \tag{2.15}$$

$$x \to -x, \phi \to -\phi, n \to -n; \tag{2.16}$$

$$y \to -y, \phi \to -\phi, n \to -n. \tag{2.17}$$

It is important that these symmetries be respected; for example, they enforce $\Gamma \to -\Gamma$ under $N' \to -N'$ in the absence of flow. We loosely enforce the symmetries by simply duplicating and transforming the training data accordingly. If desired, one may also enforce the symmetries by removing the asymmetric part of the trained DNN (we do not do this here). For example, one may symmetrize the particle flux by taking

$$\tilde{\Gamma}(\varepsilon, N', U, U', U'') = \frac{1}{4}(\Gamma(\varepsilon, N', U, U', U'') + \Gamma(\varepsilon, N', -U, U', -U'') \tag{2.18}$$

$$-\Gamma(\varepsilon, -N', U, -U', -U'') - \Gamma(\varepsilon, -N', -U, -U', -U'')). \tag{2.19}$$

Finally, it is worth mentioning that this deep learning approach has several shortcomings. For one, the assumption of space-time locality is a serious, *ad hoc* limitation. Nonlocal deep learning models may be possible, but discerning physics principles from such a model would be more challenging. The model must also be confined to a specific regime: a mean-field model will break down beyond the weak turbulence limit due to the formation of strong vortices. For this reason, we must generally choose $\kappa \lesssim 3$. A local model is also unable to capture the effect of the Kelvin-Helmholtz instability, for which the assumption of separation between mode and background scales breaks down. Another weakness is that errors are difficult to quantify meaningfully, so deciding if the DNN has been properly trained is largely a matter of physical intuition — one must check that symmetries and other constraints are respected. [That said, we have included an effort at error quantification in App. B.] Important examples of constraints

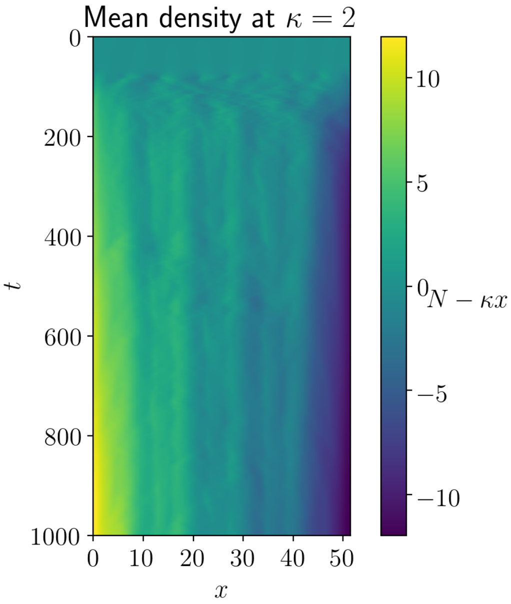


Figure 2.3: Plot of change in mean density $N - \kappa x$, where κ is the initial gradient drive, at $\kappa = 2$, from BOUT++ simulation.

which we have checked against in this work are entropy production [DII10] and the scaling of the fluxes roughly as ε^{ν} for some $\nu > 0$. Finally, the models learned by the DNN are essentially black boxes, which can reveal neither a simple mathematical function of the input parameters nor the underlying physics that led to the model. In this work, we probe the DNN models graphically and find this is sufficient to deduce the basic scalings learned by the model. More sophisticated symbolic regression approaches may be possible, but this is beyond the scope of this work.

2.3 Results

2.3.1 Numerical solution

In Figs. 2.3–2.5 we show the mean density, vorticity, and turbulent PE from a typical numerical solution of Eqs. (2.9), with gradient drive $\kappa = 2$. A few characteristics of the self-organization process are apparent. First, a small-scale ZF and a roughly uniform turbulence intensity field appear simultaneously. Then, the ZFs undergo a merger process until a large,

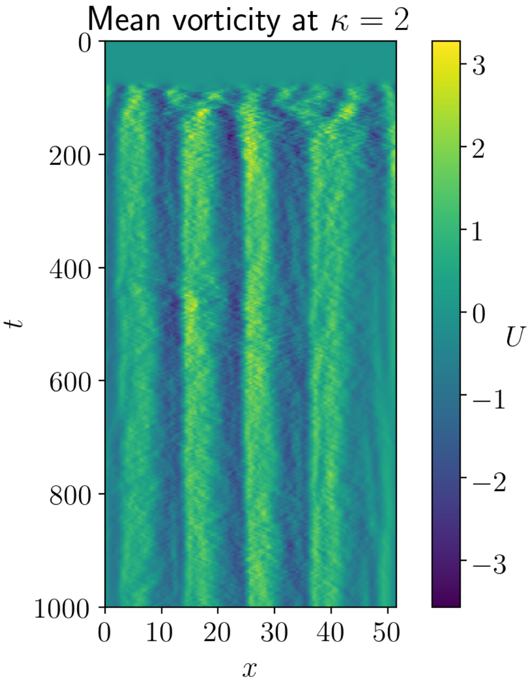


Figure 2.4: Plot of mean vorticity U at $\kappa = 2$, from BOUT++ simulation.

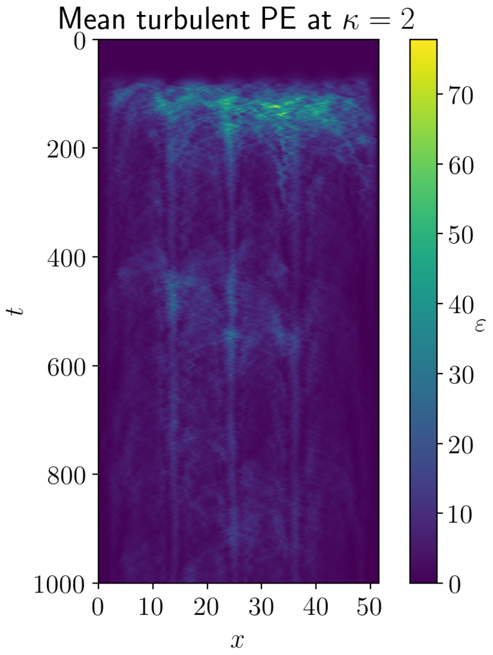
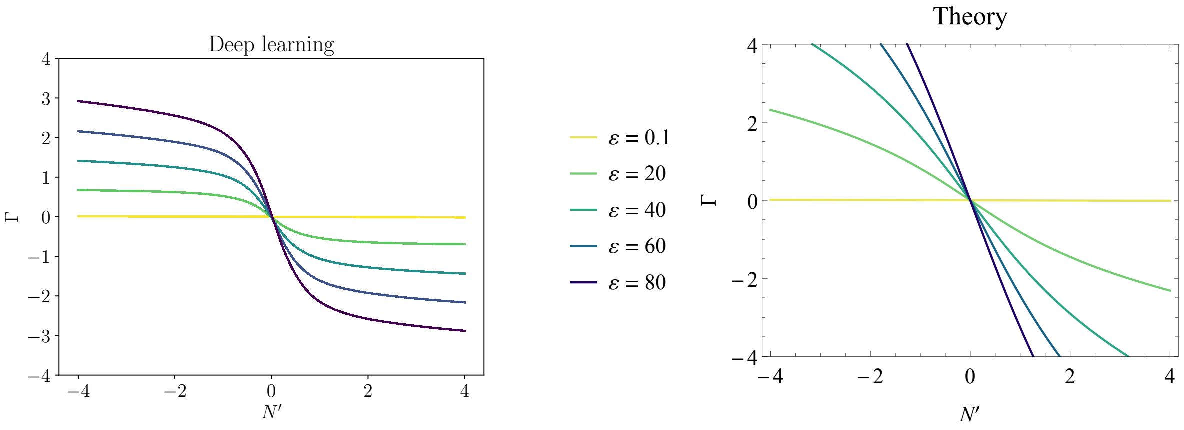


Figure 2.5: Plot of mean turbulent PE ε at $\kappa = 2$, from BOUT++ simulation. Note the appearance of corrugations near $x = 13, 25, 37$.

stable ZF scale is reached. Concurrent with the ZF evolution, the density profile is modulated, approximately in phase with the ZF. Also, particles are gradually transported down the density gradient. Meanwhile, the turbulent PE becomes concentrated at discrete corrugation sites and decays elsewhere; these sites appear to correspond with locations where U' and N' have the same sign. We will see that all of these trends are consistent with both the deep learning result and mathematical modeling.

2.3.2 Particle flux



(a)

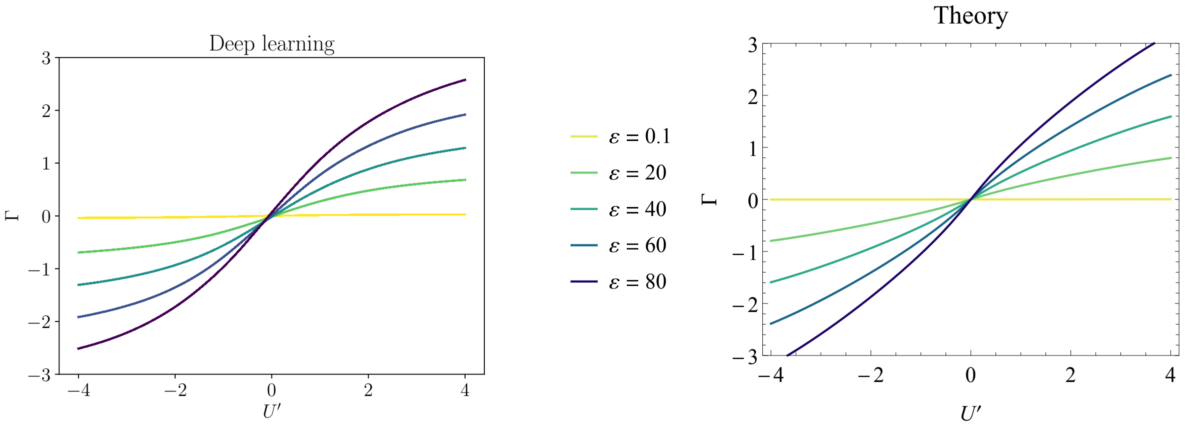
(b)

Figure 2.6: Diagonal part of the learned particle flux, at fixed $U = U' = U'' = 0$, as a function of N' and ε. The dependence on N' may be summarized as linear, plus saturation effects at large N'. The lefthand figure (adapted from Heinonen and Diamond (2020) [HD20]) shows the behavior learned by the deep neural network, whereas the righthand figure shows the prediction of the simple analytical model from Sec. 2.4 using the ansatz Lorentzian spectrum Eq. 2.33. The analytical model shows reasonable agreement when N' is small ($\lesssim 1$), but fails to accurately model saturation effects.

The DNN finds that the particle flux depends most strongly on N', U', and ε. The flux does *not* noticeably depend on U''. The basic leading-order behavior can be summarized as

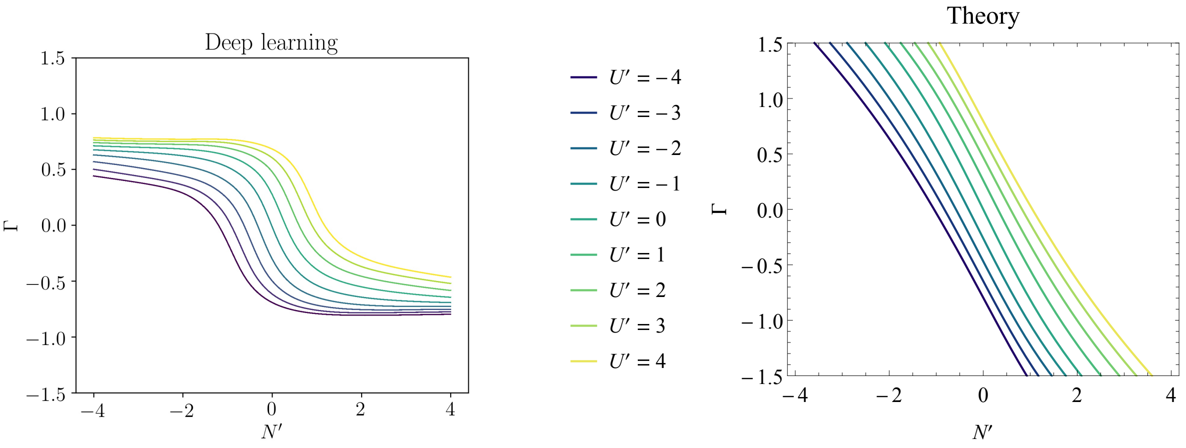
$$\Gamma \simeq \varepsilon(-D_n N' + D_u U'), \tag{2.20}$$

where $D_n \simeq 0.04$ and $D_u \simeq 0.015$. The term $\propto N'$, which we will refer to as the “diagonal” term,



(a) (b)

Figure 2.7: Same as Fig. 2.6, except for the off-diagonal part: the particle flux at fixed $N' = U = U'' = 0$, as a function of U' and ε. Again, the analytical model shows better agreement with the deep learning result when U' is small. The lefthand figure is adapted from Heinonen and Diamond (2020) [HD20].



(a) (b)

Figure 2.8: Same as Figs. 2.6–2.7, except now showing dependence on both N' and U', at fixed $U = U'' = 0$ and $\varepsilon = 20$. Near $N' = U' = 0$, the flux is roughly a linear combination of terms proportional to N' and U'. The analytical model does not well capture deviations from this behavior seen in the deep learning result. The lefthand figure is adapted from Heinonen and Diamond (2020) [HD20].

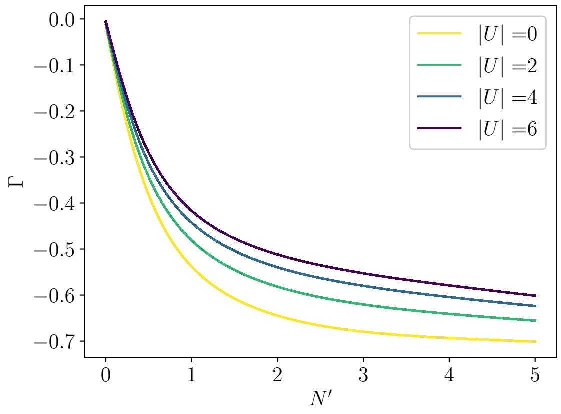


Figure 2.9: Learned particle flux at fixed $U' = U'' = 0$ and fixed $\varepsilon = 20$, as a function of N' and $|U|$ (the curves with $\pm U$ lie on top of each other). The flux is reduced by an approximate factor $(1+0.04|U|)^{-1}$.

is familiar and leads to the quasilinear relaxation of the profile. The "off-diagonal" term $\propto U'$, on the other hand, is not as well known and was first reported in Ref. [HD20]. However, the DNN indicates it is a significant effect—the flux couples to the vorticity gradient with the same order of strength as to the density gradient! We will show that the physics of this effect can be understood with a simple analytical calculation.

The off-diagonal term has immediate implications for feature formation. In the presence of a quasistable ZF, the term will tend to modulate the density profile, leading to a staircase, directly explaining the behavior of the profile presented in the previous section. More explicitly, if we set $U = U_0 \sin qx$ and fix a uniform intensity ε_0, the off-diagonal term will contribute

$$\partial_t N = -\partial_x \Gamma = D_u \varepsilon_0 U_0 q^2 \sin qx + \dots \tag{2.21}$$

to the evolution of the profile. This agrees with the observation that the density modulation tends to be in phase with that of the vorticity.

Several higher-order effects are also present. The scalings with N' and U' saturate nonlinearly and are asymptotically constant or decaying. Moreover, the gradients interact when

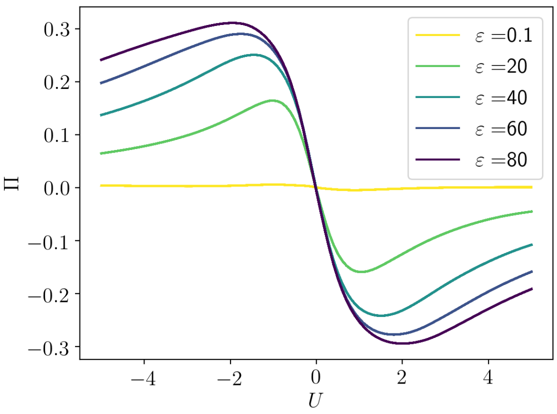


Figure 2.10: Plot of learned Reynolds stress against vorticity U at fixed $N' = 2$ and $U' = U'' = 0$ and several values of the intensity. The basic behavior is that of a negative viscosity, stabilized by nonlinear effects. From Heinonen and Diamond (2020) [HD20].

.

they become large, and can no longer be simply expressed as the simple linear combination (2.20). Notably, there is dependence on the relative sign of N' and U'. These behaviors can also be roughly explained by the same calculation, though the fine details differ.

In a final effect, which is missed by the analytical calculation in Sec. 2.4, the flux is weakly reduced (by $\lesssim 10\%$ for typical values) in the presence of a nonzero mean vorticity. Shear-induced suppression of turbulent transport is a well-known phenomenon [Ter00], but it is interesting to note that the DNN determines that the *direct* impact of the shear on the flux is weak in this system. The local *gradient* of the shear has a much stronger impact on the local flux.

All these behaviors are shown in Figs. 2.6–2.9.

2.3.3 Reynolds stress

The Reynolds stress model learned by the DNN depends strongly on U and ε, as shown in Fig. 2.10. At small U, the leading behavior is of the form

$$\Pi = \varepsilon(-\chi_1 U + \chi_3 U^3) \tag{2.22}$$

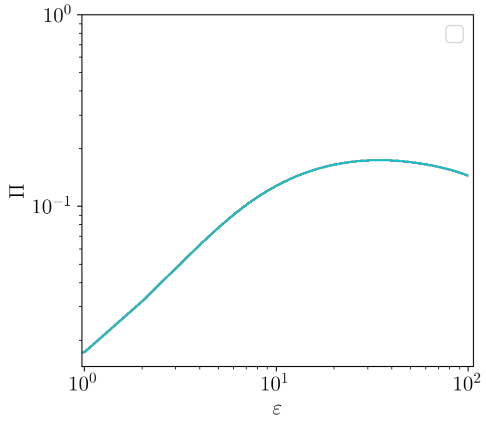


Figure 2.11: Log-log plot of learned Reynolds stress against intensity at fixed $U = 0.5$, $N' = 2$, $U' = U'' = 0$ and several values of the intensity. The scaling exponent is unity for low to moderate values of the intensity but decreases as ε increases.

.

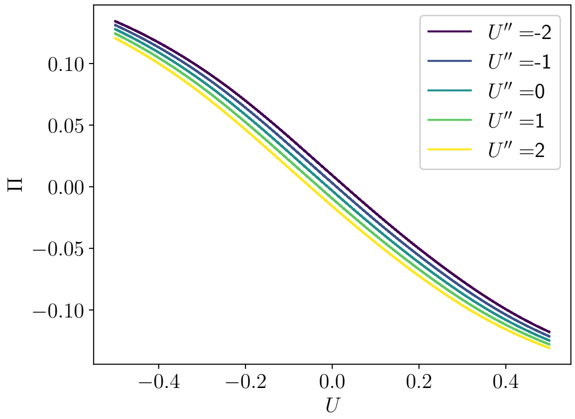


Figure 2.12: Plot of learned Reynolds stress against vorticity U at fixed $N' = 2$, $U' = 0$, $\varepsilon = 20$, and several values of U''. The leading order contribution from U'' is a stabilizing linear term. From Heinonen and Diamond (2020) [HD20].

.

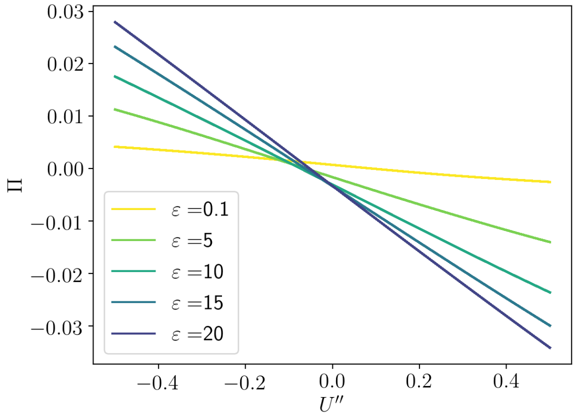


Figure 2.13: Plot of learned Reynolds stress against U'' at fixed $N' = 2$, $U = U' = 0$ and several values of the intensity. We should have $\Pi \to -\Pi$ under $U'' \to -U''$ here, but the model fails to precisely learn this, which may be attributed to the relatively small contribution to the loss function from the hyperdiffusion term. However, it is clear that this term scales roughly linearly with intensity. From Heinonen and Diamond (2020) [HD20].

.

where we estimate $\chi_1 \sim 0.015$ and $\chi_3 \sim 0.01$. Once again, there are also higher-order terms that saturate this behavior. Asymptotically, the Reynolds stress decays at large U like a power law $U^{-\nu}$, with exponent $1/2 \lesssim \nu \lesssim 1$. At high intensity, the exponent associated with the intensity scaling also saturates, approaching zero (Fig. 2.11), which is typical of strong turbulence scaling [Hor99].

The DNN also detects a stabilizing turbulent hyperdiffusion term roughly of the form $-\varepsilon\chi_4 U''$, with $\chi_4 \sim 0.0005$ (Figs. 2.12–2.13). Absent this term, the negative viscosity destabilizes all small scales, leading to unphysical blowup. It is remarkable and encouraging that the DNN is sensitive to such a small (yet important!) effect on the Reynolds stress, roughly 30 times weaker than the leading coupling to vorticity.

The above findings are consistent with higher-order quasilinear theory, from which one can obtain the model equation [DIIH05a]

$$\partial_t U = \partial_x^2(-D_1 U + D_3 U^3 - D_4 U''). \tag{2.23}$$

This equation, which might also be anticipated using Ginzburg-Landau theory, has the form of a 1-D Cahn-Hilliard equation [CH58] with dynamical coefficients, suggesting that ZF formation is associated with the spontaneous separation of positively and negatively signed vortices. The ZF grows initially due to the unstable negative viscosity term. The cubic nonlinearity stabilizes the ZF growth for large amplitudes $U \gtrsim (D_1/D_3)^{1/2}$. The hyperdiffusion D_4 originates from the leading behavior of the ZF growth rate $\gamma_{ZF} \propto q^2(1-q^2/q_0^2)$ and stabilizes small lengthscales $k \gtrsim (D_1/D_4)^{1/2}$.

The negative viscosity result also agrees with a second-order cumulant expansion (CE2) analysis of the isotropically-forced Hasegawa-Mima equation [Par15]. In contrast, in the beta-plane Navier-Stokes system (corresponding to $\rho_s \to \infty$), the ZF formation may be driven by a either a negative viscosity or negative *hyperviscosity* effect, depending on the form of the forcing [BI11, SY12].

The Reynolds stress is also found to be moderately reduced in the presence of a nonzero N' or U', by an overall factor f which behaves roughly as

$$f = \frac{1}{1 + a(N' + bU')^2} \tag{2.24}$$

with $b \approx 4$ and $a \approx 0.04$. In Fig. 2.14 we show this behavior at a fixed $U = 1$ and $\varepsilon = 20$. The expression $N' + bU'$ is the gradient of a sort of generalized vorticity, similar to the PV, except with unequal contributions from density and vorticity. This factor reduces the Reynolds stress when the vorticity gradient steepens, tending to saturate the nonlinear ZF generation and regularizing the dynamics by preventing the gradient from becoming too steep.

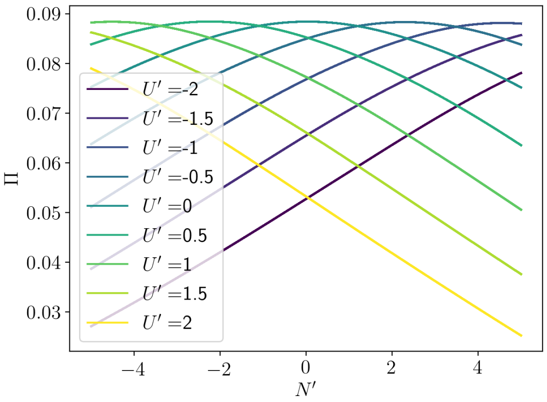


Figure 2.14: Plot of learned Reynolds stress against N' at fixed $U = 1, \varepsilon = 20$, $U'' = 0$, and several values of U'. The presence of a gradient in U' or N' tends to suppress the Reynolds stress. From Heinonen and Diamond (2020) [HD20].

.

2.4 Theory

2.4.1 Linear theory

We proceed with the linear theory of 2-D HW with a background flow. We begin with the collisionless equations

$$\partial_t \tilde{n} + N' \partial_y \tilde{\phi} + V_y \partial_y \tilde{n} = \alpha(\tilde{\phi} - \tilde{n}) \tag{2.25}$$

$$\partial_t \nabla_\perp^2 \tilde{\phi} - V_y'' \partial_y \tilde{\phi} + V_y \partial_y \nabla_\perp^2 \tilde{\phi} = \alpha(\tilde{\phi} - \tilde{n}). \tag{2.26}$$

Under the assumption of scale separation between the mean and fluctuating quantities, we can obtain the linear dispersion relation

$$(\omega - k_y V_y)^2 - \left(k_y k^{-2} V_y'' - i\alpha(1 + k^{-2})\right)(\omega - k_y V) + i\alpha k_y (N' + V_y'') = 0. \tag{2.27}$$

Separating $\omega = \omega_r + i\gamma$, one finds

$$\omega_r = k_y V_y + a/2 \pm \sqrt{\frac{1}{8}\,(a^2 - b^2 + d)} \tag{2.28}$$

$$\gamma = -b/2 \pm \sqrt{\frac{1}{8}\,(b^2 - a^2 + d)}, \tag{2.29}$$

where we have set $a = k_y V_y''/k^2$, $b = \alpha(1 + k^{-2})$, $c = \alpha k_y(N' + V_y'')/k^2$, and $d = \sqrt{(a^2 - b^2)^2 + 4(ab - 2c)^2}$.

In the adiabatic limit, one finds [2] (to leading order in $1/\alpha$) that the unstable branch has frequency

$$\omega_r = \frac{k_y(N' + V_y'')}{1 + k^2} \tag{2.30}$$

$$\gamma = \frac{k_y^2}{\alpha(1 + k^2)^3}(N' + V_y'')(k^2 N' - V_y''), \tag{2.31}$$

where we have suppressed the Doppler shift by writing $\operatorname{Re}\omega = \omega_r + k_y V_y$. There are a couple interesting points to note here. First, in the presence of a flow, the real frequency is proportional to the gradient of PV, not simply the density gradient. Moreover, there is an asymmetry to N' and V_y'' in the growth rate. A strong enough V_y'', aligned *anti-parallel* to N', will stabilize drift waves, whereas a parallel V_y'' cannot. This will have an impact on feature formation, as turbulence tends to concentrate where $V_y'' N' > 0$. This is a simple explanation for the intensity corrugations observed in the 2-D HW DNS.

---

[2] In fact, there is a second unstable branch that is driven by a strong vorticity gradient. This only occurs if, locally, $U' > |4\alpha| + N'^2/(16\alpha) + O(\alpha^{-2})$. The corresponding wavenumber is, close to threshold, $\mathbf{k} = \left(0, 1 - |N'|/(8\alpha) + O(\alpha^{-2})\right)$. This mode appears to be exotic and of questionable relevance to the present work.

2.4.2 Quasilinear fluxes and wave-kinetic equation

We can now compute the quasilinear particle flux in the adiabatic limit:

$$\begin{aligned}\Gamma &= \mathrm{Re}\int d^2\mathbf{k} - ik_y\tilde{n}_{\mathbf{k}}\tilde{\phi}^*_{\mathbf{k}} \\ &= \int d^2\mathbf{k}\,\frac{-k_y^2N'(\gamma_{\mathbf{k}}+\alpha)+\alpha k_y\omega_{\mathbf{k}}}{\omega_{\mathbf{k}}^2+(\gamma_{\mathbf{k}}+\alpha)^2}|\tilde{\phi}_{\mathbf{k}}|^2 \\ &\simeq -\frac{1}{\alpha}\int d^2\mathbf{k}\,\frac{k_y^2}{1+k^2}\left(k^2N'-U'\right)|\tilde{\phi}_{\mathbf{k}}|^2 \qquad (2.32)\end{aligned}$$

where we have dropped the subscript r from ω and again written $U = V_y'$. Thus to leading order, the flux indeed separates into a linear combination of diagonal and off-diagonal terms. Estimating $k^2 = 2$ in accordance with the most unstable mode (when $U' = 0$) having $\mathbf{k} = (0, \pm\sqrt{2})$, the coupling to vorticity gradient is roughly half that to density gradient. Both of these results agree with the deep learning result. In Figs. 2.6–2.8, we plot the analytical result (2.32) using the exact frequencies (2.28)–(2.29) and an ansatz Lorentzian spectrum centered about the most unstable mode

$$\varepsilon_{\mathbf{k}} = \frac{\varepsilon}{2\pi^2\Delta k_x\Delta k_y}\frac{1}{1+k_x^2/\Delta k_x^2}\left(\frac{1}{1+(k_y-\sqrt{2})^2/\Delta k_y^2}+\frac{1}{1+(k_y+\sqrt{2})^2/\Delta k_y^2}\right), \qquad (2.33)$$

where we have set $\Delta k_x = \Delta k_y = 0.8$. The normalization has been chosen so that $\int d^2\mathbf{k}\,\varepsilon_{\mathbf{k}} = \varepsilon$. As compared to the DNN result, we see that the theory captures well the behavior at small N' and U', but the agreement is poor when either gradient, especially N', is large.

The quasilinear Reynolds stress is given by

$$\Pi = \int d^2\mathbf{k}\,k_xk_y|\tilde{\phi}_{\mathbf{k}}|^2. \qquad (2.34)$$

Additional physics input is needed to obtain a mean-field model; we defer to the next subsection.

To close the mean-field dynamics, one needs an evolution equation for the turbulence intensity. One can use $\varepsilon_{\mathbf{k}} \simeq (1+k^2)^2|\tilde{\phi}_{\mathbf{k}}|^2$ and perform an asymptotic expansion in $1/\alpha$ and the zonal flow scale q [SD99] to obtain the wave-kinetic equation (WKE)

$$\partial_t \varepsilon_{\mathbf{k}} + \partial_{k_x}\omega_{\mathbf{k}}\partial_x\varepsilon_{\mathbf{k}} - (\partial_x\omega_{\mathbf{k}} + k_y U)\,\partial_{k_x}\varepsilon_{\mathbf{k}} = (2\gamma_{\mathbf{k}} + \partial^2_{k_x,x}\omega_{\mathbf{k}})\varepsilon_{\mathbf{k}}. \tag{2.35}$$

The unusual term $\partial^2_{k_x,x}\omega_{\mathbf{k}}$ comes from the non-Hermiticity of the time evolution operator in the fluctuation equations (2.25–2.26) and can be derived with the Wigner-Moyal formalism [RPSD16]. This term breaks drift wave quanta conservation but is necessary to preserve conservation of the total PE, so that, upon integrating over **k**-space, Eq. (2.35) is consistent with the equation for the mean turbulent PE

$$\partial_t \varepsilon + 2(\Gamma - \partial_x \Pi)(N' + U') = 0. \tag{2.36}$$

Equation (2.36) results from subtracting (2.26) from (2.25), multiplying both sides by the fluctuating PV $\tilde{n} - \nabla^2\tilde{\phi}$, averaging, and neglecting the turbulent PE flux $\langle(\tilde{n} - \nabla^2\tilde{\phi})^2\tilde{v}_x\rangle$. The turbulent PE flux gives rise to turbulence spreading, which is neglected in this study.

2.4.3 Reynolds stress

We now attempt to model the learned behavior of the Reynolds stress. Using the wave-kinetic equation, consider the response to a finite background shear U and a uniform density gradient N'. Neglect contributions from U'' as higher-order in q, and also assume U' is small enough that the response is uniform and the group velocity term can be neglected. Using the method of characteristics and neglecting the evolution of U, one sees the solution to the WKE will lie on curves parameterized as

$$k_x = k_{x0} - k_y U t. \tag{2.37}$$

Such a use of the shearing coordinates of Goldreich and Lynden-Bell [GLB65] follows,

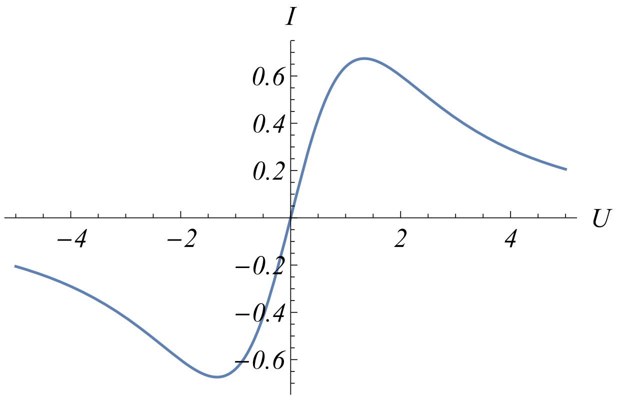


Figure 2.15: Plot of integral I appearing in Eq. (2.39), for $N' = 2, U' = 0, \alpha = 2, \tau = 0.5$. Compared to Fig. 2.10, the Reynolds stress is overestimated by an order of magnitude.

for example, Kim and Diamond (1999) [KD03]. Taking $k_{x0} = 0$, we then have a solution

$$\varepsilon_{\mathbf{k}}(t) = \varepsilon \exp\left(2\gamma_k|_{k_x=-k_yU t}t\right). \tag{2.38}$$

Replacing t with a correlation time τ, the evolution equation for the vorticity is then

$$\begin{aligned}\partial_t U = -\partial_x^2 \Bigg[\varepsilon \int dk_y \, & \frac{k_y^2 U\tau}{(1+k_y^2(1+U^2\tau^2))^2} \\ & \times \exp\left(\frac{2k_y^2\tau}{\alpha(1+(1+U^2\tau^2)k_y^2)^3}(N'+U')(k_y^2(1+U^2\tau^2)N'-U') \right) \Bigg].\end{aligned} \tag{2.39}$$

The integral (call it I) is plotted for $N' = 2, U' = 0, \alpha = 2, \tau = 0.5$ in Fig. 2.15. In the limit $U \to \infty$ one has the asymptotic behavior

$$I \simeq \frac{\pi U\tau}{2(1+U^2\tau^2)^{3/2}} \sim U^{-2}. \tag{2.40}$$

In the opposite limit $U \to 0$, we have

$$I \simeq aU\tau - bU^3\tau^3 \tag{2.41}$$

where a and b are integrals which depend on U', N', α, τ. For N' and U' small we have

$$a \simeq \frac{\pi}{2}\left(1+\frac{(N'+U')(5N'-3U')\tau}{32\alpha}\right) \tag{2.42}$$

and

$$b \simeq \pi\left(\frac{3}{4}+\frac{5(N'+U')(5N'-3U')\tau}{128\alpha}\right). \tag{2.43}$$

Up to the hyperdiffusion, which is neglected in this calculation, this limit is in basic qualitative agreement with the DNN results. It is worth noting, however, that Eq. 2.39 is the result of numerous severe approximations and, compared with the DNN model, overestimates the Reynolds stress by an order of magnitude. Notably, we have overlooked the fact that in a real system, the exponential growth of the turbulence will quickly saturate due to nonlinear effects. However, this calculation still captures three essential qualitative features of its U dependence: (a) negative viscosity for small U, stabilized by (b) a cubic nonlinearity (both of which depend weakly on the gradients), and (c) power-law decay at large U.

2.4.4 Reduced 1-D model

The WKE, together with the evolution equations

$$\partial_t N = \partial_x \int d^2\mathbf{k} \frac{k_y^2}{\alpha(1+k^2)^3}(k^2 N' - U')\varepsilon_{\mathbf{k}} \tag{2.44}$$

$$\partial_t U = \partial_x^2 \int d^2\mathbf{k} \frac{k_x k_y}{(1+k^2)^2}\varepsilon_{\mathbf{k}}, \tag{2.45}$$

represent a closed system for the mean field dynamics which conserves total PE. To make direct contact with the deep learning model, we must remove dependence on the spectrum and demote $\varepsilon_{\mathbf{k}}$ to ε. In making such an approximation, temporal memory effects as well as scale dependence of the correlation time are lost, but the dynamics become far easier to discern.

Using the DNN result as inspiration, we first use a simple model for the particle flux consisting of the diagonal and off-diagonal terms and a weak nonlinear saturation effect,

$$\Gamma = \frac{-D_n \varepsilon N' + D_u \varepsilon U'}{1 + c_1 N'^2 + c_2 U'^2} \tag{2.46}$$

with $D_n = 4, D_u = 1.5, c_1 = c_2 = 0.05$. The parameters are chosen for rough consistency with the DNN result, with time sped up by a factor of 100. For the Reynolds stress, we set

$$\Pi = -\frac{c_3 \varepsilon}{1 + a(N' + bU')^2} \frac{\tanh U/U_0}{(1 + c_4 U^2)^{1/2}} - \chi_4 \varepsilon U'', \tag{2.47}$$

with $c_3 = 1.5, U_0 = c_4 = 1, \chi_4 = 0.05, a = 0.04, b = 4$. This form is chosen for having the correct asymptotic behavior for large and small U. The factor $f = 1/(1 + a(N' + bU')^2)$ was included because it was found to be beneficial to the stability of the numerical solution by smoothing the vorticity profile. Absent this factor, kinks tend to form in U, leading to a breakdown in the numerical solution.

We then close the dynamics using conservation of turbulent PE, Eq. (2.36), whence we have the system

$$\partial_t N = \partial_x \left(\frac{D_n \varepsilon \partial_x N - D_u \varepsilon \partial_x U}{1 + c_1 (\partial_x N)^2 + c_2 (\partial_x U)^2} \right) - D_0 \partial_x^4 N \tag{2.48}$$

$$\partial_t U = \partial_x^2 \left(-\frac{c_3 \varepsilon}{1 + a(N' + bU')^2} \frac{\tanh U/U_0}{(1 + c_4 U^2)^{1/2}} - \chi_4 \varepsilon \partial_x^2 U \right) - \mu U - D_0 \partial_x^4 U \tag{2.49}$$

$$\begin{aligned} \partial_t \varepsilon = & \left[\frac{D_n \varepsilon \partial_x N - D_u \varepsilon \partial_x U}{1 + c_1 (\partial_x N)^2 + c_2 (\partial_x U)^2} \right. \\ & \left. + \partial_x \left(-\frac{c_3 \varepsilon}{1 + a(N' + bU')^2} \frac{\tanh U/U_0}{(1 + c_4 U^2)^{1/2}} - \chi_4 \varepsilon \partial_x^2 U \right) \right] (\partial_x N + \partial_x U) \\ & - \gamma_d \varepsilon - \gamma_{NL} \varepsilon^2 + D_\varepsilon \partial_x^2 \varepsilon. \end{aligned} \tag{2.50}$$

We include a linear turbulence damping $\gamma_d = 0.3$ which sets a threshold gradient drive $\kappa_0 = \sqrt{\gamma_d / D_n}$ for turbulence growth, and a nonlinear damping $\gamma_{NL} = 0.1$ which saturates the turbulence growth at a finite level and represents nonlinear transfer to dissipation. The nonlinear term is explicitly neglected in the quasilinear approximation and must be included *ad hoc*; the exponent two is consistent with expectations from weak turbulence theory. γ_d was chosen for consistency with the chosen 2-D HW DNS parameters (Sec. II-A), for which the linear stability threshold is $\kappa_0 \simeq 0.286$ (as may be computed from a slight modification of Eq. (2.27). Flow damping $\mu = 1$, also consistent with the DNS parameters, was included as well. However, at this value of μ, the damping had little interesting effect on the dynamics. (Hyper-)diffusion terms $D_0 = D_\varepsilon = 0.01$ are also included to improve the stability properties of the system. We initialize N with a uniform gradient

$$N(x,t=0) = \kappa x, \tag{2.51}$$

U with a small inhomogeneity

$$U(x,t=0) = 0.001 \sin\frac{6\pi x}{L}, \tag{2.52}$$

and ε with a small uniform intensity

$$\varepsilon(x,t=0) = \varepsilon_0 \tag{2.53}$$

with $\varepsilon_0 = 0.001$. We employ the boundary conditions

$$N'(x=0,t) = N'(x=L,t) = \kappa, \tag{2.54}$$

$$N(x=0,t) = 0, \tag{2.55}$$

$$N(x=L,t) = \kappa L, \tag{2.56}$$

$$\begin{aligned} U(x=0,t) &= U''(x=0,t) \\ &= U(x=L,t) = U''(x=L,t) = 0, \end{aligned} \tag{2.57}$$

$$\varepsilon(x=0,t) = \varepsilon(x=L,t) = \varepsilon_0. \tag{2.58}$$

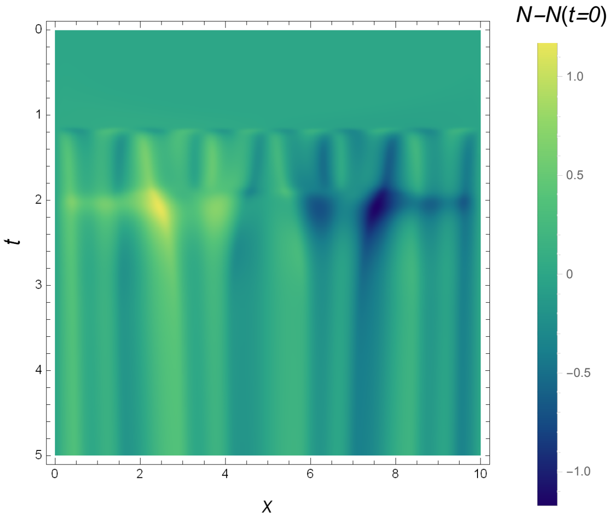


Figure 2.16: Color map showing solution $N - N(t = 0)$ to 1-D model with gradient drive $\kappa = 1.5$.

Numerically, this highly stiff system presents multiple challenges, including the presence of the small lengthscale $(\chi_4/\chi_1)^{1/2}$ which must be resolved and the formation of sharp corrugations in ε with steep gradients. Moreover, blowup rapidly occurs if ε is allowed to go spuriously negative anywhere. We solve it, for $\kappa = 1.5$, on a box of size $L = 10$ (corresponding to $\rho_* = 1/10$) using the implicit Lobatto-IIIC Runge-Kutta algorithm (which is well-suited for stiff problems) of order 4 [Jay15] with grid spacing $\Delta x = 0.01$.

The solutions are shown in Figs. 2.16–2.18. We have checked that they properly conserve U, N, and the total PE to good approximation.

There are three basic stages of evolution, illustrated by time slices in Figs. 2.19–2.21. First, the turbulence field grows uniformly in response to the driving gradient. When the turbulence field is large enough, it induces a turbulent Reynolds stress that spontaneously drives a ZF. The ZF, in turn, induces a staircase pattern in the density profile via the off-diagonal particle flux, as well as corrugations in the turbulence intensity which are due to modulation of the turbulence growth rate — indeed, the corrugations are localized where U' is parallel to N', increasing the growth rate.

Finally, the vorticity field tilts in response to the corrugation of the intensity profile. This

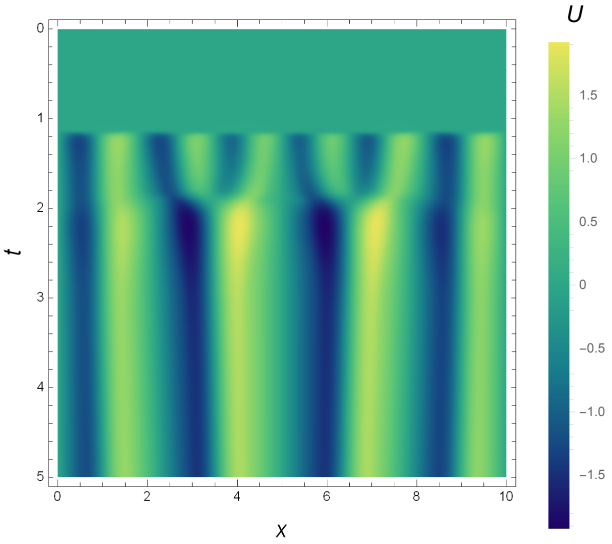


Figure 2.17: Color map showing solution U to 1-D model with gradient drive $\kappa = 1.5$.

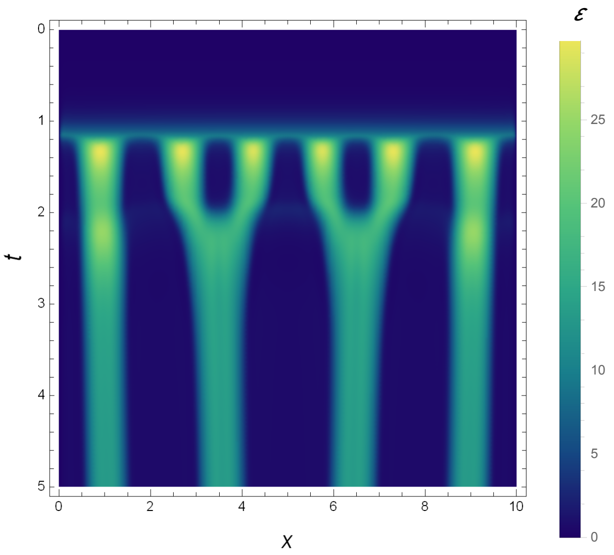


Figure 2.18: Color map showing solution ε to 1-D model with gradient drive $\kappa = 1.5$.

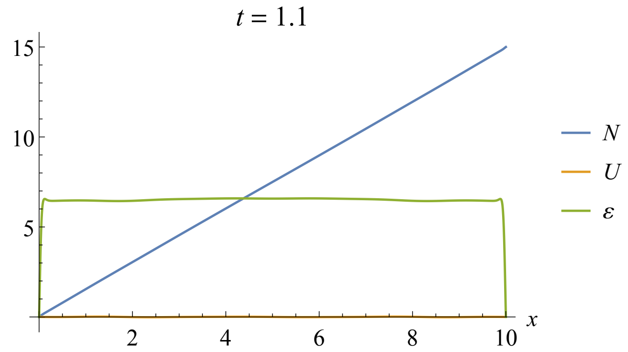


Figure 2.19: Plot of solution to 1-D model at $t = 1.1$, illustrating first stage of evolution. The turbulence field grows uniformly with the density gradient as the free energy source. No zonal flow has developed.

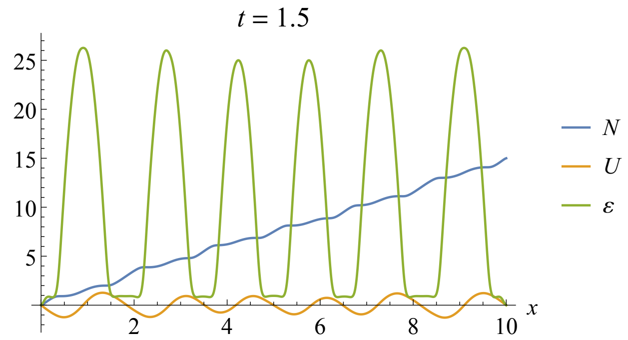


Figure 2.20: Plot of solution to 1-D model at $t = 1.5$, illustrating second stage of evolution. A zonal flow spontaneously forms, corrugating the intensity profile and inducing a staircase in the density profile.

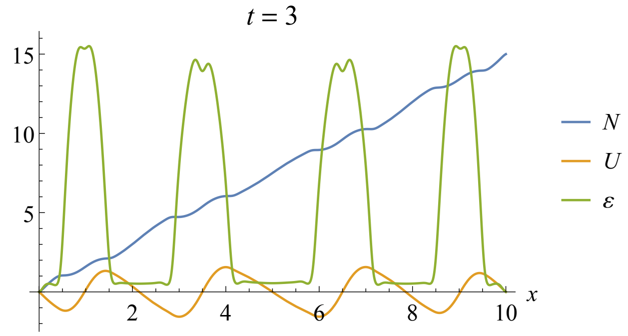


Figure 2.21: Plot of solution to 1-D model at $t = 3$, illustrating final stage of evolution. The zonal flow field tilts in response to the intensity corrugation and merges into a quasisteady, persistent flow.

occurs because the intensity peaks result in a greater local Reynolds stress, which in turn locally steepens the vorticity gradient via negative viscosity. Thus a pattern of alternating steep and not-steep U' emerges. This tilting effect is also seen in the 2-D HW DNS (Sec. II-A). The tilting is symptomatic of the breaking of radial reflection symmetry by the density gradient.

As the ZF tilts, zonal layers merge, leading to the formation of a quasisteady profile. Intensity corrugations also merge. Compared to the 2-D HW DNS, this merger process is simpler and contains fewer stages. This may be due to a finite size effect, as our simulation box for the 1-D model was smaller. It is also possible that some physics of the merger process is lost when approximating the local spectrum $\varepsilon_{\mathbf{k}}$ by the local intensity ε, as the spectrum is likely to differ at early times.

2.4.5 Comparison to Ashourvan-Diamond model

Ashourvan and Diamond proposed [AD17, AD16] an analytic model for staircasing and feature formation with a bistable mixing length ansatz. Similar to our model, it self-consistently evolves mean density, mean vorticity, and mean turbulent PE while conserving the total PE. The key structure-forming physics input is the ansatz for the mixing length, which is linked to the physics of the Rhines mechanism [Rhi75] (in the spirit of Balmforth *et al.* [BSY98]):

$$\ell_{mix}^2 = \frac{\ell_0^2}{\left(1+\frac{\ell_0^2(N'+V_y'')^2}{\varepsilon}\right)^{\kappa}}. \tag{2.59}$$

For example, the particle flux in this model is given by $\Gamma = -c\ell_{mix}^2 \varepsilon N'$. This choice of mixing length, a hybrid of a forcing scale ℓ_0 and a Rhines scale $\ell_{Rh} = \sqrt{\varepsilon}/|N' + V_y''|$, models the inhomogeneous mixing of potential vorticity (PV). The Rhines scale is defined by the crossover of the eddy turnover rate, which is of the order $\varepsilon^{1/2}$, and the three-drift-wave mismatch frequency $\omega_{MM} = \omega_{\mathbf{k}} - \omega_{\mathbf{k}'} - \omega_{\mathbf{k}-\mathbf{k}'} \sim \omega_{\mathbf{k}}$. At scales exceeding the Rhines scale, turbulence is wavelike,

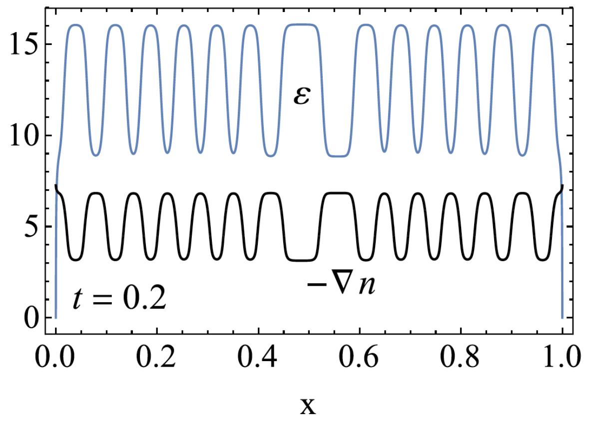


Figure 2.22: Quasiperiodic features in the density and turbulent PE profiles in the AD model. This figure appeared previously in Ashourvan and Diamond (2017) [AD17].

and at shorter scales it is eddy-like. The assumption motivating the mixing-length ansatz is that the Rhines scale should be the dominant spatial scale for the turbulence when the PV gradient is strong, and the forcing scale should dominate when it is weak. (See Ref.[GDH$^+$19] for a related model where the mixing length is based upon a correlation time associated with shearing, as well as a comparison of the outputs of these two mixing-length ansatzes.)

This model generates staircases in the density profile by a mechanism which is associated with the fact that the flux is bistable with respect to the driving gradient. The mechanism is a feedback loop wherein the steepening of the PV gradient reduces the local flux, which further enhances the gradient, etc.

It is worth comparing and contrasting staircase formation in the present model versus in that of Ashourvan-Diamond (AD) (see Fig. 2.22). For one, there is a difference in shape: in the AD model, the jumps in N' are sharper, due to the nonlinear self-focusing effect of the mixing length. Such an effect is absent in the relatively simpler off-diagonal flux. A second observation is that in the AD model, N and U tend to have a relative phase of π, whereas in the present model, they tend to be in phase. Finally, in the present model, the modulation of N is slaved to the modulation of U, whereas in the AD model, a staircase would still form in the absence of any ZF.

2.5 Discussion

We have used a new deep learning-based approach to probe turbulence dynamics in the HW system and build a reduced model. Using the new method, we have explicitly verified previous Cahn-Hilliard-like models for spontaneous flow generation via the Reynolds stress. Moreover, our results have highlighted a previously unreported off-diagonal particle flux which couples to the ZF, and shown that this effect leads to staircase formation. This off-diagonal flux is a consequence of the nonlinear convection of vorticity, which induces a shift in the drift-wave frequency. The deep learning method picks this effect out as important, especially relative to the direct effect of the local shear. Finally, we have shown, via numerical solution, that the detailed reduced model inferred by the deep learning method is reasonable and consistent with direct numerical simulation of the full 2-D HW system.

The staircasing effect induced by the off-diagonal flux may be understood as a new feedback loop in the drift-wave/ZF system. The profile drives the turbulence via linear instability. The turbulence, in concert with a small seed inhomegeneity, gives rise to a Reynolds stress, producing a quasiperiodic ZF pattern. Finally, the ZF feeds back on the profile by modulating the particle flux, steepening the profile in some places and flattening it in others.

The 2-D HW system to which we have applied our method is especially simple and has a number of useful symmetries. It is natural to ask to what extent our method is portable to other applications — to more complicated models, or even experiments, where data may be harder to obtain and there are fewer symmetries or analytical results available to guide us. While the structure and hyperparameters of the DNN will inevitably require tuning from problem to problem, we speculate that training a reduced model with our method is likely to be successful, generally speaking, if three criteria are satisfied: (a) we can identify important mean-field variables on which the fluxes likely depend, (b) there is a minimal degree of symmetry in the problem, and (c) sufficient training data are available.

Criterion (a) is necessary for feature selection; we must be able to define and identify the independent mean-field variables in order to define our model. In most systems, typical candidates are profile gradients, flow velocity and/or shear, the magnetic field strength, and the turbulence intensity. However, if too many such variables are deemed important, or if they are too strongly coupled to be considered independent, it may be difficult to make sense of or adjudicate the correctness of any model that is trained.

Criterion (b) is a basic necessity for our method to make physical sense. In essence, this method consists of a data-driven reduction of dimensionality. While one can envision generalizations which coarse-grain the system in other ways, any such reduction can only be valid if the system possesses enough symmetry, either globally or locally, to motivate the coarse-graining. Here, the method succeeds because of approximate poloidal symmetry, but it ceases to make sense when the turbulence becomes strong enough that vortex interactions dominate and break this symmetry. Equivalently, there is a threshold in turbulent enstrophy density beyond which our method is inapplicable; this corresponds to the breakdown of weak turbulence theory. One can reasonably extrapolate and expect this method might work best in an H-mode regime where axisymmetric flows are present.

Finally, criterion (c) ensures that training can actually converge. In particular, the data should span a sufficiently large portion of parameter space, which in practice requires performing many runs, with initial conditions cleverly and efficiently tuned from run to run. The need for many runs, in turn, demands computational speed. 2-D HW is especially fast for the purposes of data generation, but we are confident that simple models in three dimensions (3-D HW being an obvious example) can be also solved numerically fast enough on a high-performance machine. It is less clear that our method will work for complex gyrokinetic codes, which may take millions of core-hours for a single run, but we take heart that the data generation scales very well with the system size and simulation time. Finally, while nothing in principal precludes the use of experimental data for this method, the need for a considerable degree of resolution in both time and space may render it impractical.

It is also important to note that any exact symmetries which can be identified are extremely beneficial, if not necessarily crucial, for training. In the present work, these aided in virtually all aspects of the process: feature selection, data generation, and verification that the learned model is physically reasonable.

As previously discussed, the new method makes the explicit assumption of space-time locality, which is quite severe and cannot be rigorously justified. In fact, there is considerable evidence that non-local processes have important effects on the turbulent dynamics [ISS$^+$15, DPDG$^+$10]. Our method selects, in principle, the local mean-field model that can *best* explain the dynamics, but some physics is almost certainly lost. We note that nonlocal generalizations of our approach may be possible; for instance, the imposition of spatial locality might be relaxed by designing a deep learning model where the entire radial density, flow, and intensity profiles are treated as input variables. On the other hand, such a model would be more challenging to interpret.

It is possible that other nonparametric methods, such as local regression or spline methods, could be effectively utilized in place of an MLP for the present application. This could be explored in another study.

Future work will focus on studying such generalizations of our approach, applying our method to more complicated systems as well as the problem of turbulence spreading, and further studying the impact of the off-diagonal flux on transport and feature formation.

Chapter 2, in full, is a reprint of the material as it appears in R. A. Heinonen and P. H. Diamond, *Plasma Phys. Control. Fusion* 62, 105017 (2020). The dissertation author was the primary investigator and author of this paper.

Chapter 3

Subcritical turbulence spreading and avalanche birth

3.1 Introduction

A challenge of modeling transport in magnetic fusion plasmas is that the local fluctuation level is not determined solely by the temperature and density profiles. In general, there is a *nonlocal* relation between fluxes and profiles. The degree or range of the nonlocality, an important object of study, depends on the precise mechanism controlling the nonlocal dynamics. The radial propagation of turbulent fluctuations, referred to as *turbulence spreading*, is one such mechanism, since the turbulence can easily propagate beyond the microturbulence correlation length [ISS$^+$15]. As a result, the fluctuation level (and thus fluxes) will generally have a nonlocal, non-Fickian dependence on the driving gradient. Explicitly, the local, Fickian flux-gradient relation $Q(r) = -\chi \nabla T$ will be altered to an integral relation of the form

$$Q(r) = -\chi \int dr' K(r,r') \nabla T(r'), \tag{3.1}$$

with the kernel K representing a generalized diffusivity, with variable mesoscopic range. The physics of the breakdown of Fick's law due to mesoscopic transport events is reviewed by Hahm and Diamond (2018) [HD18].

The spreading of turbulence is a mesoscale phenomenon, meaning it occurs on lengthscales much larger than the turbulence correlation length or eddy size but much smaller than the minor radius. The relevant timescales of spreading are slower than the turbulence correlation time but faster than transport timescales.

The consequences of spreading in confined plasmas are numerous. Fluctuations can penetrate into the linearly stable regions of the plasma and excite turbulence there [HDL$^+$04, NNR05]. Spreading is believed to be involved in the observed breakdown of gyro-Bohm transport scaling [LH04]. Spreading is also closely related conceptually to avalanching, a process wherein a large initially localized excitation cascades into neighboring regions, exciting turbulence as in a domino effect and causing strong transport events or bursts. Avalanching, a feature of systems that exhibit self-organized criticality, is especially likely in marginal or weakly subcritical systems [SG98].

An interesting feature of turbulence spreading is that, despite its delocalizing effect, it can itself be described using local dynamics. The simplest approach is to employ a 1-D, nonlinear reaction–diffusion model for the turbulence intensity, of the form

$$\partial_t I = f(I) + \partial_x (D(I) \partial_x I), \tag{3.2}$$

where $f(I)$ describes the total local nonlinear growth/decay of the turbulence intensity, and $D(I)$ represents nonlinear diffusion of the turbulence energy. The nonlinear diffusion originates from nonlinear mode coupling in wavenumber space, which results in spatial scattering. Of these models, the most common is based on the Fisher equation [HDL$^+$04, NNR05, GD05]. This model, which takes f to be quadratic, i.e.

$$f(I) = \gamma_0 I - \gamma_{nl} I^2, \tag{3.3}$$

incorporates effects of local linear turbulence drive and local nonlinear coupling to dissipation that saturates the linear drive. The Fisher model predicts ballistically propagating fronts of turbulence, provided the system is supercritical, i.e. has positive linear growth. In particular, turbulence propagates superdiffusively, with $\Delta x \sim t$, rather than $\Delta x^2 \sim t$.

The Fisher model succeeds in explaining — at a fundamental level — why the local flux-gradient relation is violated. It also reveals a superdiffusive character of turbulence spreading and predicts a radial propagation speed of the order $c \sim \sqrt{D_0 \gamma_0}$, where γ_0 is the local linear growth rate and D_0 is the nonlinear turbulence diffusion coefficient. These predictions are in rough agreement with simulation [SII$^+$12].

However, under careful scrutiny, the Fisher model may be seen to suffer from several important limitations. First, it does not predict robust turbulence spreading in a marginal or subcritical state, and for this reason cannot be used as a model for avalanching. Instead, spreading only occurs in the supercritical regime. In a system with fully developed turbulence, such a description of turbulent front propagation is unphysical, since in a linearly unstable zone, noise will have excited the system to a saturated turbulence level in the first place. Moreover, even during the initial growth of the fluctuation envelope, Fisher-type spreading can only be relevant if the rate of propagation exceeds the linear growth rate $\gamma_0 \ll c/\Delta x$, i.e. $\Delta x^2 \gamma_{nl} \ll D_0$, where Δx is an appropriate mesoscopic length scale.

The Fisher model also predicts very weak penetration into a linearly stable zone—indeed, to a finite depth on the order of just a few correlation lengths [GD05]. This calls into question the utility of the Fisher model for explaining any observed nonlocality in the turbulence intensity or fluctuations in a stable zone.

As a final reason to reconsider the Fisher model, subcritically unstable turbulence has been observed in a variety of contexts in magnetic fusion (MF) plasma simulations. This phenomenon cannot be captured in the supercritical Fisher model. For example, magnetic shear has a stabilizing effect on drift eigenmodes, but simulations of the 3-D Hasegawa–Wakatani system suggest that turbulence can self-sustain even when all modes are stable [BW85]. The underlying nonlinear stability mechanism [DZB95] can be traced to spontaneous sheared radial flows, which set up poloidally-varying density perturbations. These density perturbations, in turn, act as a free energy source for drift waves with small k_y which are immune to magnetic shear damping. Finally, the drift waves amplify the original radial flows, completing the feedback loop. Self-sustaining, self-organized turbulence has also been observed in simulations of more a complicated sheared system based on the Braginskii equations [Sco90, Sco92]. Finally, gyrokinetic simulations with a strong ambient perpendicular shear flow imposed indicate that the turbulence in such a system is subcritical, as well [BPH$^+$11, vWHS$^+$16].

These considerations motivate the introduction of a new model for the spreading of turbulence. In this paper, we analyze the simplest extension of the Fisher model that can incorporate subcritical turbulence — or more generally, bistability. It is of the form of Eq. (3.2), with the *cubic* reaction function

$$f(I) = \gamma_1 I + \gamma_2 I^2 - \gamma_3 I^3. \tag{3.4}$$

This model is closely related to the FitzHugh–Nagumo model for excitable media [Fit61, NAY62]. While the underlying physics of the bistability is not our focus here, we note that Guo and Diamond (2017) [GD17] showed that temperature profile corrugations can provide an additional nonlinear drive corresponding to the quadratic term in Eq. (3.4) that results in essentially the same model as the one we consider in this paper.

Temperature profile corrugations are a feature of $E \times B$ staircases, which are mesoscale structures that have been observed to spontaneously self-organize in tokamak plasmas [DPDG$^+$10]. Staircases consist of a quasi-periodic pattern of $E \times B$ shear layers equipped with corrugations, with regions of strong avalanche activity in between. Guo and Diamond (2017) showed that the global effect of the staircase pattern is bistability in the turbulence intensity. This bistability may result in a hysteresis when the system is subjected to an external flux (see Sec. 3.3.3); the branches of this hysteresis are determined by global features of the profiles. This differs from the hysteresis associated with a localized internal transport barrier, whose branches are determined by local variables and their gradients. As we will argue, the same threshold behavior that induces global hysteresis results *locally* in a size threshold for slugs of turbulence, which in turn helps explain the ubiquitous presence of avalanches in the steps of the $E \times B$ staircase.

We will show that the bistable model has several attractive features, especially relative to the Fisher model. In addition to addressing subcritical turbulence, it also predicts significantly stronger penetration of turbulence into a linearly stable zone, thus achieving stronger nonlocality. The additional nonlinear drive results in propagation of the turbulence intensity at constant speed through the stable zone, rather than exponential decay over a finite depth, as long as the linear damping is sufficiently weak. Moreover, the model predicts turbulence propagation in the relevant marginal and weakly subcritical regimes, in addition to the obvious case of supercriticality. As mentioned above, there is a (double) threshold — in intensity and spatial size — for the spreading of an initially localized slug of turbulence in a subcritical system. We derive the size threshold by a simple but novel physical argument, based on a competition between local growth and outward diffusion of turbulence in the slug. We argue that this threshold behavior signals that sufficiently strong, intermittent local fluctuations can propagate and excite turbulence on much larger scales. This captures an essential feature of avalanching.

In Sec. 3.2, the Fisher model is reviewed and its key results are presented. In Sec. 3.3.1, the bistable model is introduced and the underlying physics is discussed. In Secs. 3.3.2–3.3.4

the model is analyzed; parameter regimes, stable points, free energy, hysteresis, and traveling wavefront solutions are discussed. In Sec. 3.3.5, the threshold for spreading of a slug is derived and implications are discussed. Finally, in Sec. 3.4, the propagation of turbulence into a bistable system with weak damping is considered, and it is shown that the propagation is ballistic.

3.2 Fisher model

We begin by introducing and reviewing the relatively familiar unistable model for turbulence spreading, which is a reaction–diffusion equation with a stabilizing nonlinearity and nonlinear diffusion. Explicitly, the model equation is

$$\partial_t I = \gamma_0 I - \gamma_{nl} I^{1+\beta} + \partial_x (D(I)\partial_x I), \tag{3.5}$$

where I is the turbulence intensity and $D(I) = D_0 I^\beta$, with $\beta \simeq 1/2$ characteristic of strong turbulence and $\beta \simeq 1$ characteristic of weak turbulence [Hor99]. Throughout this work, we will take $\beta = 1$ so that the nonlinearity in I is quadratic and $D(I) = D_0 I$. γ_0 and γ_{nl} are, respectively, linear and nonlinear growth rates. (We note in passing the possibility of a convective term $v_g \partial_x I$ due to group velocity effects, but this can be removed by a Galilean transformation to the comoving frame [HDL$^+$04]. However, if the system is inhomogeneous, the coefficients in Eq. (3.5) become time-dependent under that transformation; this may have unexplored consequences, for example, in the system considered in Sec. 3.3.6.)

The first term on the RHS ($\propto I$) represents local linear turbulence growth or damping, the second ($\propto I^{1+\beta}$) represents local nonlinear saturation, and the third represents nonlinear diffusion of the turbulence energy. γ_{nl} includes contributions from both zonal flow shear and mode–mode coupling effects. The nonlinear diffusion term arises as a consequence of nonlinear $E \times B$ coupling of modes, by using a closure of the $E \times B$ convective nonlinearity [KDM$^+$03] that treats the small modulation scale and large envelope scale on an equal footing:

$$\sum_{\mathbf{k}'}(\mathbf{k}\cdot\mathbf{k}'\times\hat{z})^2|\tilde{\phi}_{\mathbf{k}'}|^2 R(\mathbf{k},\mathbf{k}')I_{\mathbf{k}} \to -\frac{\partial}{\partial x}D_x(I)\frac{\partial}{\partial x}I_{\mathbf{k}} + \mathbf{kk}:\mathbf{D}I_{\mathbf{k}}, \tag{3.6}$$

where R is a resonance function that determines the correlation time, $D_x = \sum_{\mathbf{k}'} k_y'^2 |\tilde{\phi}_{\mathbf{k}'}|^2 R(\mathbf{k},\mathbf{k}')$, and $\mathbf{D} = \sum_{\mathbf{k}'} \mathrm{diag}(k_x'^2, k_y'^2)|\tilde{\phi}_{\mathbf{k}'}|^2 R(\mathbf{k},\mathbf{k}')$. The first term is the radial nonlinear diffusion and the second represents local nonlinear transfer.

Up to the nonlinearity in the diffusion, Eq. (3.5) is known as the Fisher–KPP (for Kolmogorov, Petrovskii, and Piscounov) equation, or simply the Fisher equation, and is used in a number of applications including chemistry and population dynamics [Mur02]. Equation (3.5) can also be derived via a Fokker–Planck formalism, in which case, the nonlinear diffusion results from a random walk in intensity [GD05]. It is worth noting that such an approach requires the local transition pdf (i.e. the probability for the intensity to take a “step” of a certain size) to have a finite second moment; relaxing this assumption would demand a fractional kinetic model.

When $\gamma_0 > 0$, Eq. (3.5) has an unstable fixed point at $I = 0$, representing a state of low or vanishing turbulence, and a stable fixed point at $I = \gamma_0/\gamma_{nl}$, representing a saturated state of finite turbulence. The latter corresponds to the standard mixing-length estimate for the saturation level. If there is instead linear damping ($\gamma_0 < 0$), the turbulent root becomes negative and unphysical, while the laminar root at $I = 0$ becomes stable. The basic dynamics of this system, illustrated in Fig. 3.1, are as follows: When $\gamma_0 > 0$, perturbations develop into asymptotic traveling waves connecting the two roots, which spread outward at constant speed [Aro80, New80]

$$c = \sqrt{\frac{D_0\gamma_0{}^2}{2\gamma_{nl}}}. \tag{3.7}$$

Note that the nonlinear diffusion contributes a steepening term $\sim(\partial_x I)^2$ and a smoothing term $\sim I\partial_x^2 I$. Since the latter vanishes when $I = 0$, the fronts formed are sharp, with a discontinuity in the spatial derivative of the intensity profile at $I = 0$.

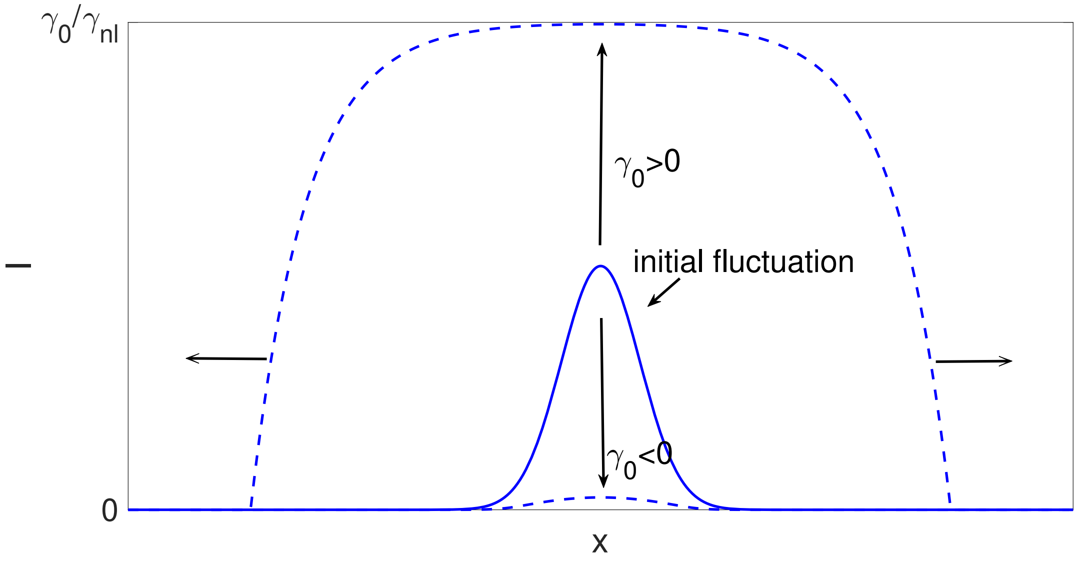


Figure 3.1: An initial fluctuation in the Fisher model will grow into a wave and spread if $\gamma_0 > 0$ or decay to 0 if $\gamma_0 < 0$.

Thus, we have ballistic turbulence spreading in the supercritical regime. If instead $\gamma_0 < 0$, any perturbation from $I = 0$ — a turbulent fluctuation, for instance — will decay to linear order like $e^{-|\gamma_0|t}$.

This exponential decay of turbulence in the subcritical regime suggests that, according to the Fisher model, turbulence is strongly damped in the stable zone. One can verify this by considering the spreading of turbulence from a linearly unstable zone to a linearly stable zone. This can be modeled with the system

$$\begin{cases} \partial_t I = \gamma_g I - \gamma_{nl} I^2 + \partial_x (D_0 I \partial_x I), & x < 0 \\ \partial_t I = -\gamma_d I - \gamma_{nl} I^2 + \partial_x (D_0 I \partial_x I), & x > 0. \end{cases} \tag{3.8}$$

The fixed points of the model can be solved for analytically. A traveling wave of turbulence propagating from the left will penetrate a finite depth λ into the stable region, forming a stationary, exponentially decaying intensity profile (Fig. 3.2). The penetration depth is on the order of the

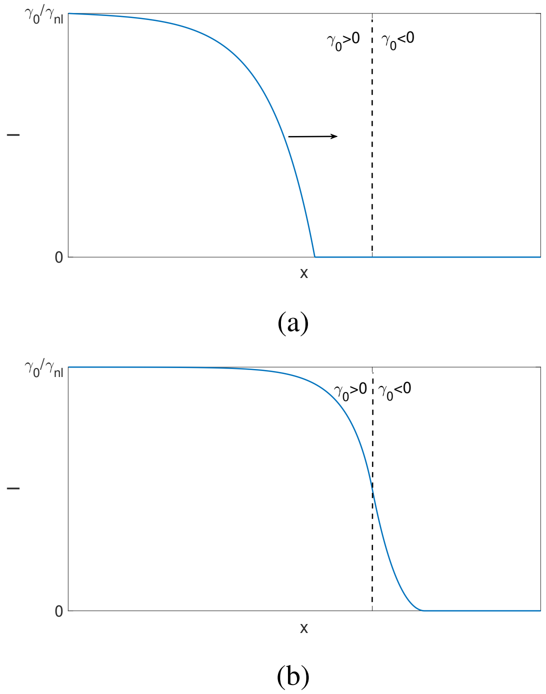


Figure 3.2: A Fisher wave front develops in the unstable zone (a) and penetrates a short depth into the stable zone, asymptotically forming a stationary intensity profile (b).

turbulence mixing length,

$$\lambda \sim \sqrt{\frac{D_0}{\gamma_{nl}}} + O\left(\log \frac{\gamma_g}{\gamma_d}\right), \tag{3.9}$$

assuming $\gamma_g \gtrsim \gamma_d$ [GD05]. For typical gyro-Bohm diffusive scaling of D_0 and γ_{nl}, one then expects λ_0 to be on the order of a few ρ_i. This is very weak penetration, and it calls into question the efficacy of turbulence spreading in a Fisher model as a mechanism of delocalization, especially when one considers clear experimental evidence of fluctuations even in stable regions of the plasma [NSK$^+$05]. This observation, coupled with the need for a model that can support subcritical turbulence, motivates the search for a new model.

3.3 Bistable model

3.3.1 Introduction of the model

We now replace the Fisher model with a bistable one. The simplest generalization is of the form

$$\partial_t I = f(I) + \partial_x(D(I)\partial_x I), \tag{3.10}$$

where we again take $D(I) = D_0 I$, and the reaction function $f(I)$ is now bistable in some parameter regime. The lowest order polynomial that meets this criterion is cubic, so we take

$$f(I) = \gamma_1 I + \gamma_2 I^2 - \gamma_3 I^3, \tag{3.11}$$

where γ_2 and γ_3 are assumed to be positive. This is similar to the Fisher model in that it includes linear turbulence drive and nonlinear saturation (the cubic term), but it also includes an intermediate nonlinear instability term ($\propto I^2$). (It is worth noting that the cubic nonlinearity is just a model intended to encapsulate any and all nonlinear damping effects, including transfer to dissipation, resonance effects, etc.) This is by no means the only choice for f, but it is attractive for its simplicity—it is the minimal model for subcritical turbulence spreading. It is also closely related to the well-studied FitzHugh–Nagumo equations originally used to model neuron activation. In this paper, the model is used as generic and phenomenological, as it captures the basic dynamics of any bistable turbulence spreading system. Using typical drift-wave scalings, one anticipates the γ_i to be roughly on the order of the diamagnetic frequency ω_* and the diffusivity D_0 to be on the order of the gyro-Bohm diffusivity $\chi_{GB} \sim \rho_i^2 c_s/a$.

Alternatively, we can arrive at the same model by way of a more concrete, theoretical approach. Compare Eqs. (3.10) and (3.11) with the intensity equation derived by Guo and

Diamond [GD17]:

$$\partial_t I = \gamma_0(\langle A\rangle - A_C)I + \gamma_0\langle A\rangle\left(\frac{D_0}{\chi_{neo}+|\chi_m|}\right)^{1/2} I^{3/2} - \gamma_{nl}I^2 + \partial_x\left(D_1 I\partial_x I\right). \tag{3.12}$$

First, note that there is no essential difference in form between these equations. For example, taking $I \rightarrow I^2$ in Eq. (3.12) results in a nearly identical cubic equation, differing only by unimportant $O(1)$ factors leading the terms contributed by the diffusion. The first term on the RHS of Eq. (3.12) is the linear gradient drive under the critical gradient hypothesis: A denotes $-\partial_x T$ so that $\langle A\rangle$ and A_C are the mean and critical temperature gradients, respectively. The third term is the usual nonlinear saturation, with γ_{nl} the nonlinear damping rate. The fourth term, again, represents nonlinear diffusion of turbulence due to mode coupling, so that $D_1 I$ is the nonlinear intensity diffusivity. The terms listed thus far correspond closely to those in the Fisher model equation.

The crucial second term, which gives rise to bistability, is an effect due to mesoscopic corrugations in the temperature profile, which have been observed in gyrokinetic simulations [WABC06]. One way these corrugations can be induced is by inhomogeneous turbulent mixing, which occurs due to reduced turbulent heat flux in the zonal flow shear layer. In turn, this forces heat to accumulate near the shear layer. Equivalent to this accumulation, there is a mesocale upgradient heat flux $\tilde{Q}_{T,m} = \chi_m \tilde{A}_m$, where $\tilde{Q}_{T,m}$ is the mesoscale turbulent heat flux, χ_m is a *negative* thermal conductivity, and $\tilde{A}_m$ is the mesoscale temperature gradient fluctuation. The local accumulation of heat steepens the local temperature gradient and, near marginal stability, can cause it to exceed the critical gradient (Fig. 3.3). This, in turn, nonlinearly drives growth of the turbulence intensity. The nonlinearity results because the turbulence itself drives fluctuations in the mean square temperature gradient, creating a feedback. In particular, we have the Zel'dovich relation

$$\langle \tilde{A}_m^2\rangle = \frac{D_0\langle A\rangle^2}{\chi_{neo}+|\chi_m|} I, \tag{3.13}$$

where $\tilde{A}_m$ is the mesosopic temperature gradient fluctuation, D_0 is the turbulent conductivity (so

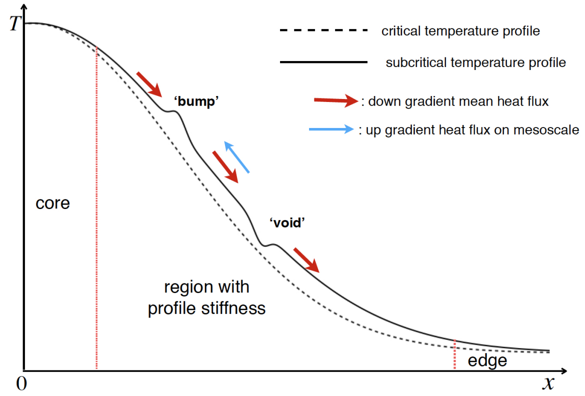


Figure 3.3: Profile corrugations ('bumps' or 'voids') can cause the critical gradient to be locally exceeded. This leads to an additional nonlinear mesoscale turbulence drive and causes the turbulence intensity to be bistable.

that the mean turbulent heat flux is $\langle Q_T \rangle = D_0 I \langle A \rangle)$, and χ_{neo} is the neoclassical heat conductivity. This last result can be obtained by multiplying the temperature evolution equation

$$\partial_t T + \nabla \cdot Q_T = \partial_x (\chi_{neo} \partial_x T) + S\delta(x) \tag{3.14}$$

(here S is the heating source) by T, averaging over micro-timescales and mesoscales, and substituting the above expressions for $\langle Q_T \rangle$ and $\tilde{Q}_{T,m}$.

It is also worth noting the similarity of our model to some models for the onset of turbulence in pipe flow [BSM$^+$15, Pom15], as well as the Landau–Ginzburg model of Gil and Sornette (1995) for avalanching in sandpiles [GS96]. The latter model includes an order parameter S that controls the activity of sand grains and a local sandpile height h, whose dynamics are described by

$$\partial_t S = \gamma \left(\left(\frac{(\partial_x h)^2}{(\partial_x h_c)^2} - 1 \right) S + \alpha S^3 - S^5 \right) \tag{3.15}$$

$$\partial_t h = D \partial_x (S^2 \partial_x h), \tag{3.16}$$

where γ, α, D are parameters and $\partial_x h_c$ is the critical gradient beyond which grains begin to topple.

S, which also obeys a bistable dynamical equation, thus plays the role of the turbulence field I. $\partial_x h$, which is analogous to a driving temperature (or density) gradient, controls the linear drive. The dynamics of h resemble that of a turbulent heat equation. Including nonlinear diffusion of the field S would then yield a simple extension of our model that includes evolution of the local gradient by turbulent diffusion of heat. Such an extension obviously reduces to our model in the limit when the heat diffusion timescale is much longer than the timescales associated with the S dynamics, i.e. $D \ll \gamma/(\partial_x h_c)^2$ and $D \ll D_0$, where D_0 is the nonlinear turbulence intensity diffusivity. Less obvious is the fact that it essentially reduces to our model in the opposite limit. See Appendices C–D for further discussion).

3.3.2 Fixed points and free energy

We now proceed with an analysis of the model. Consider the uniform fixed points of Eq. (3.10), i.e. the roots of the polynomial f. Assuming $\gamma_2{}^2 > -4\gamma_1\gamma_3$, there is one root at $I = 0$ and a pair of roots at $I = I_\pm = (\gamma_2 \pm \sqrt{\gamma_2{}^2 + 4\gamma_1\gamma_3})/2\gamma_3$. The root at $I = I_+$, representing a saturated "high turbulence" state, is always stable. Note that it is nonzero, i.e. the system can be excited to turbulence, even at marginal linear stability $\gamma_1 = 0$.

If the system is linearly damped ($\gamma_1 < 0$), the root at $I = 0$ ("low turbulence") is stable and the root at $I = I_-$ is unstable. If, on the other hand, there is linear growth ($\gamma_1 > 0$), the root at $I = 0$ becomes unstable, and the root at $I = I_-$ becomes negative and is no longer physical. The latter scenario is qualitatively similar to an unstable Fisher system, while the former is bistable and exhibits different dynamics. We focus our attention on the bistable case; see Table 1 for a summary of the various parameter regimes.

Table 3.1: Summary of features of the various parameter regimes in the cubic model. Unphysical roots ($I < 0$) are ignored.

Regime	Stable roots	Unstable roots	Fronts?	Comments
$\gamma_1 > 0$	I_+	0	forward-propagating	similar to Fisher with $\gamma_0 > 0$
$\gamma_1 < 0$, $\lvert\gamma_1\rvert\gamma_3/\gamma_2^2 < \frac{15}{64}$	$0, I_+$	I_-	forward-propagating	$\alpha < \alpha^*$; turbulent root abs. stable
$\gamma_1 < 0$, $\frac{15}{64} < \lvert\gamma_1\rvert\gamma_3/\gamma_2^2 < \frac{1}{4}$	$0, I_+$	I_-	receding	$\alpha > \alpha^*$; turbulent root metastable
$\gamma_1 < 0$, $\lvert\gamma_1\rvert\gamma_3/\gamma_2^2 > \frac{1}{4}$	0	none	none	similar to Fisher with $\gamma_0 < 0$

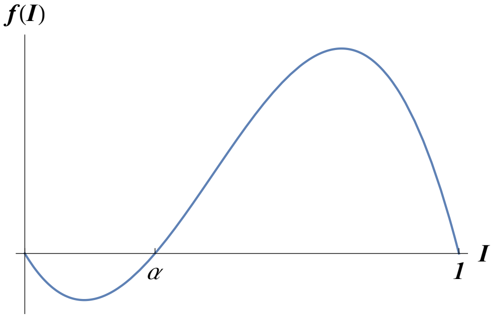


Figure 3.4: Typical plot of the cubic reaction function $f(I)$ in the bistable regime

Transforming $I \to I_+ I$ and redefining

$$|\gamma_3| {I_+}^2 \to \gamma, \tag{3.17}$$

$$\frac{I_-}{I_+} \to \alpha, \tag{3.18}$$

$$I_+ D_0 \to D \tag{3.19}$$

yields the simplified equation

$$\partial_t I = \tilde{f}(I) + \partial_x(\tilde{D}(I)\partial_x I), \tag{3.20}$$

where $\tilde{f}(I) = \gamma I(1-I)(I-\alpha)$ and $\tilde{D}(I) = DI$ (we will drop the tildes henceforth). This bistable, rescaled version of $f(I)$ is plotted in Fig. 3.4. The remaining dimensional parameters γ and D can also be scaled out of the problem via the transformation $x \to (D/\gamma)^{1/2}x$ and $t \to t/\gamma$, which provides the natural length and time scales of the model.

The unstable root α, which lies in the open interval $(0,1)$, increases with γ_1, and γ_3 and decreases with γ_2. Generally speaking, α measures the stability of the laminar root. In fact, when

α exceeds a critical value α^*, the system transitions from a weakly stable to a strongly stable state. To understand this, it helps to rewrite the model equation in variational form

$$D(I)\partial_t I = -\frac{\delta \mathscr{F}}{\delta I} \tag{3.21}$$

with Lyapunov functional

$$\mathscr{F}[I] \equiv \int dx \left[\frac{1}{2}(D(I)\partial_x I)^2 - \int_0^I dI' \, D(I') f(I')\right]. \tag{3.22}$$

We have by an elementary calculation that

$$\frac{d\mathscr{F}}{dt} \leq 0 \tag{3.23}$$

at all times.

$\mathscr{F}$ may be identified as a kind of free energy associated with the turbulence field, the dissipation of which governs the nonequilibrium dynamics. The first term in the integrand is a "kinetic" term due to the flux of turbulence intensity $D(I)\partial_x I$, and the second a "potential" term due to the local growth of the turbulence intensity. The free energy is minimized by gradient descent, at a local rate given by the inverse of the nonlinear diffusion $D(I)$. Equation (3.21) can also be seen as an infinite-degrees-of-freedom generalization of the equation of motion

$$D_{\alpha\beta}\frac{\partial x_\beta}{\partial t} = -\frac{\partial V}{\partial x_\alpha} \tag{3.24}$$

for an inertia-free particle in a potential $V(\mathbf{x})$ with (diagonal) viscous damping tensor $D_{\alpha\beta}$.

The potential

$$V(I) = -\int_0^I dI' \, D(I') f(I') \tag{3.25}$$

has minima at $I = 0, 1$ and a maximum at $I = \alpha$. The value of α governs which of the minima

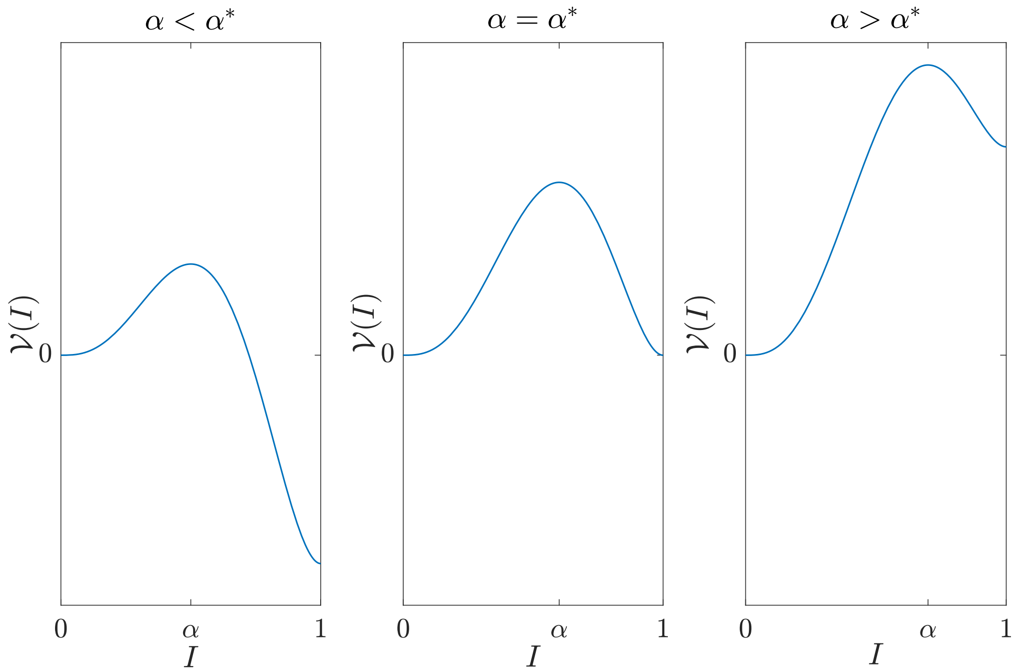


Figure 3.5: Behavior of potential $V(I)$ for various α and uniform I. Below α^*, the laminar state $I = 0$ is only metastable as the true minimum is at the turbulent state $I = 1$. Above α^*, $I = 1$ is metastable and $I = 0$ is absolutely stable.

is absolutely stable: below $\alpha^* = 3/5$, the laminar root is metastable and the turbulent root is absolutely stable, while above α^* the reverse is true (see Fig. 3.5).

3.3.3 Hysteresis

The potential barrier at the unstable root $I = \alpha$ leads to a local threshold for the spreading of a turbulent slug (see Sec. 3.3.5). There is also a global threshold, which can result in hysteresis when the system is subjected to an external source/sink. In the laminar state, applying a heat flux larger than a forward threshold will induce a transition to the turbulent state. Likewise, in the turbulent state, reducing the flux below the backward threshold will induce a transition back to the laminar state. These thresholds are generally unequal, leading to hysteresis. The fluctuation level I corresponds to a mean global heat flux $|Q| = D_0 I |\langle \partial_x T \rangle|$, where here $D_0 \sim \chi_{GB}$ is the turbulent thermal diffusivity. Neglecting the neoclassical flux, the forward threshold occurs at the unstable root

$$|Q_{for}| = D_0 I_- |\langle \partial_x T \rangle|. \tag{3.26}$$

When this threshold is exceeded, the mean flux will evolve to a steady state $|Q| = D_0 I_+ |\langle \partial_x T \rangle|$. Thus, to induce a back-transition, the flux must be reduced by a value

$$|Q_{back}| = D_0 (I_+ - I_-) |\langle \partial_x T \rangle|. \tag{3.27}$$

The strength of the hysteresis is then computed as

$$\frac{|Q_{back}| - |Q_{for}|}{|Q_{for}|} = 1 - 2\frac{I_-}{I_+} = 1 - 2\alpha. \tag{3.28}$$

Curiously, if $0.5 < \alpha < 0.6$, the hysteresis has negative strength — the flux required to induce a back-transition is weaker than that to induce a forward transition — despite the turbulent root being absolutely stable. This does not occur in the case of linear diffusion, in which case $\alpha^* = 0.5$.

Employing the Guo–Diamond model yields, in the weakly subcritical regime,

$$I_+ \simeq \frac{\gamma_0^2 \langle A \rangle^2}{\gamma_{nl}^2} \frac{D_0}{\chi_{neo} + |\chi_m|} \tag{3.29}$$

$$I_- \simeq \varepsilon \frac{\chi_{neo} + |\chi_m|}{D_0}, \tag{3.30}$$

where $\varepsilon \equiv 1 - \frac{\langle A \rangle}{A_c}$. The hysteresis strength near marginality is then $1 - 2\varepsilon \left(\frac{\gamma_0 \langle A \rangle}{\gamma_{nl}} \frac{D_0}{\chi_{neo} + |\chi_m|} \right)^2$.

The transition at α^* is reminiscent of a barrier transition, in that it signals a reduction of turbulent transport in the system and can be induced by $E \times B$ shear flow, which contributes to the linear damping in the system.

Experiments of Inagaki *et al.* [ITT$^+$13] on the Large Helical Device purported to demonstrate hysteresis in both the fluctuation intensity and heat flux with respect to the local temperature gradient (Fig. 3.6). The claimed hysteresis was global and occurred in the absence of an internal transport barrier; this would suggest a bistable S-curve relation between the gradient and the turbulence intensity and directly support our model. However, in our view, the claim of hysteresis was insufficiently substantiated, and the results can be well-explained by an instantaneous dependence of the intensity on the input power.

We strongly suggest further experiments in the vein of Inagaki *et al.* (2013) which more clearly demonstrate and characterize the turbulence bistability. Greater resolution of the dependence of the fluctuation intensity on the input power is necessary to determine if there is any threshold behavior in the system, and a more careful study of the relaxation after the ECH is turned off would help clarify the stability of the state excited by the ECH. Information about the fluctuation spectrum and spatial correlations and simultaneous measurements of the zonal flow pattern would also be very interesting for determining how the system evolves as we move around the hysteresis loop.

3.3.4 Traveling fronts

Similar to the Fisher model, in the supercritical case $\gamma_1 > 0$, there are traveling, constant speed wavefront solutions in this model which connect the laminar and turbulent roots. However, unlike in the Fisher model, such waves also exist in the case $\gamma_1 < 0$, as long as the system is bistable (i.e. with sufficiently weak damping) [SGM97]. That is, this model permits *subcritical* spreading of turbulence. This is more physically realistic, for in a supercritical zone and in the presence of any noise, one expects the turbulence to be in a saturated state to begin with.

The waves propagate with constant speed of order $c \sim \sqrt{D\gamma} = \sqrt{I_+{}^3 D_0 \gamma_3}$, as can be surmised by dimensional analysis. Near marginal linear damping, we have $c \sim \sqrt{D_0 \tilde{\gamma}}$, where $\tilde{\gamma} = \gamma_2{}^3/\gamma_3{}^2$.

The sign of the speed determines whether turbulence spreads or recedes. It can be found by writing $I(x,t) = I(z)$, where $z = x - ct$, multiplying Eq. (3.20) by $D(I)I'(z)$, and integrating,

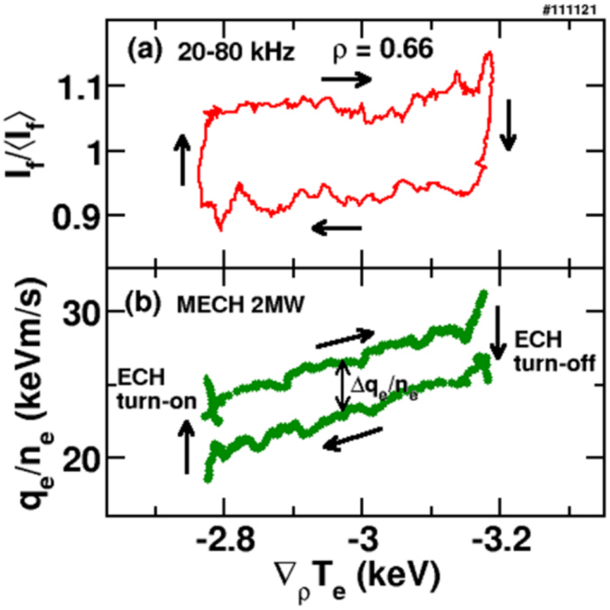


Figure 3.6: Experiments of Inagaki *et al.* (2013) purported to demonstrate hysteresis in the fluctuation intensity and heat flux with respect to the temperature gradient. Pictured are the intensity (a) and flux (b) as a function of the gradient, subject to an ECH source which is turned on and then off. Reproduced with permission from *Nuclear Fusion* **53**, 113006 (2013). Copyright 2013 IAEA, Vienna.

which shows that

$$c\int_{-\infty}^{\infty} D(I(z))I'(z)^2\,dz = \int_0^1 D(I)f(I)\,dI, \tag{3.31}$$

and thus the sign of c is given by the sign of the integral on the RHS. In the linear diffusion case ($D(I) = \text{const}$), this becomes an equal-area rule for f, so the propagation direction is given by a kind of generalized Maxwell construction. This integral is (up to a sign) the same as the potential identified in Sec. 3.3.2, and one finds that the speed switches sign when $\alpha = \alpha^*$, in agreement with the intuition developed in the free energy picture.

An exact expression for the wave speed does not appear to be known, but for $\alpha < \alpha^*$ (positive wave speed) there is an excellent variational principle-based approximation due to Pedersen [Ped05] (following Mornev [Mor98]), which gives (in units of $\sqrt{D\gamma}$)

$$c \approx -\frac{\lambda V(1)}{\sqrt{2\langle V\rangle - V(1)}} = \lambda\left(\frac{1}{10}-\frac{\alpha}{6}\right)\left(\frac{15}{1-\alpha}\right)^{1/2}, \tag{3.32}$$

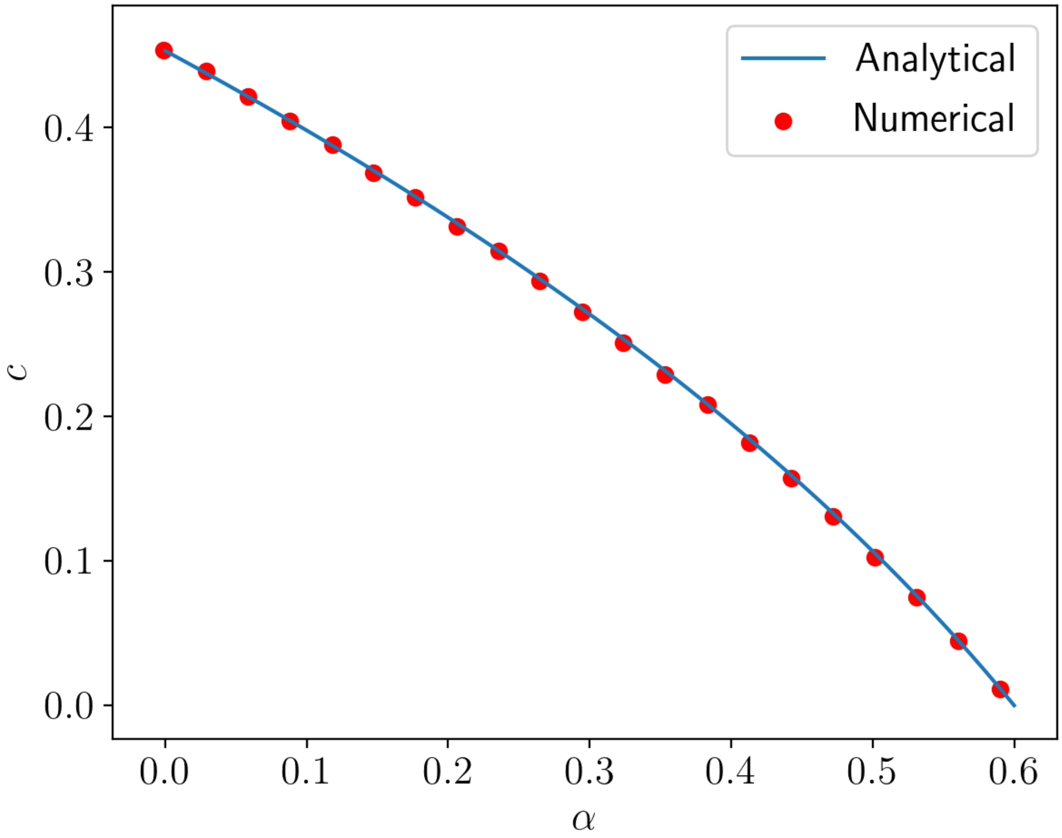
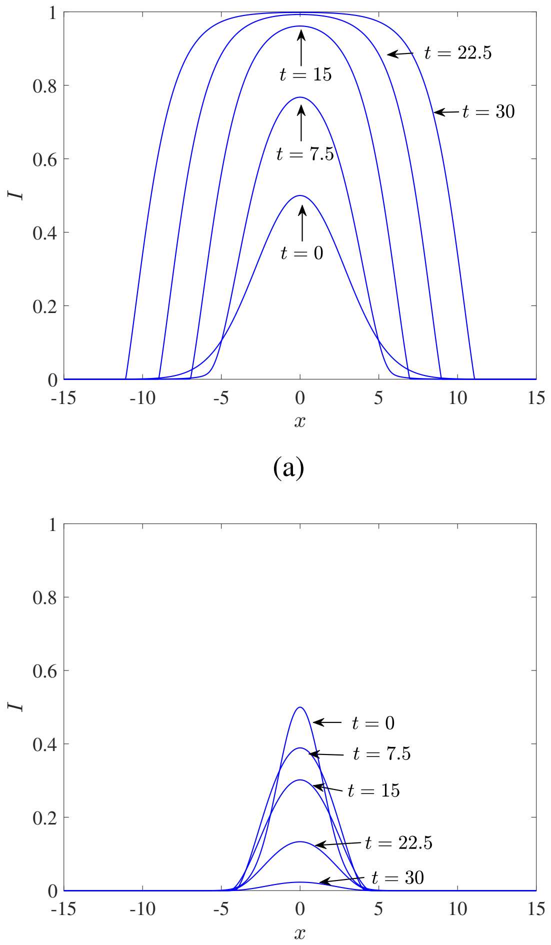


Figure 3.7: Numerically obtained front propagation speed c as function of unstable root α (data from [Ped05]), compared with Mornev–Pedersen analytical approximation (3.32), for $\alpha < \alpha^*$.

where $\lambda \approx 1.17$ is a chosen $O(1)$ factor and $\langle V \rangle = \int_0^1 dI\, V(I)$. This approximation is compared to numerical results for $\alpha < \alpha^*$ in Fig. 3.7.

3.3.5 Threshold for spreading

A classic question in turbulence theory, important to the larger problem of the transition from laminarity to turbulence, asks under what conditions an initially localized spot or slug of turbulence will spread. In this bistable model, assuming $\alpha < \alpha^*$, a slug of turbulence (i.e. some localized perturbation from $I = 0$) must exceed simultaneous amplitude and spatial size thresholds



(a)

(b)

Figure 3.8: A slug will either grow into a wave (a) or collapse (b) depending on its "size." These numerical solutions, which differ only in the width of the (Gaussian) initial condition, were obtained at $\alpha = 0.3$ using the MATLAB routine `pdepe`. x is in units of D/γ and t is in units of γ^{-1}.

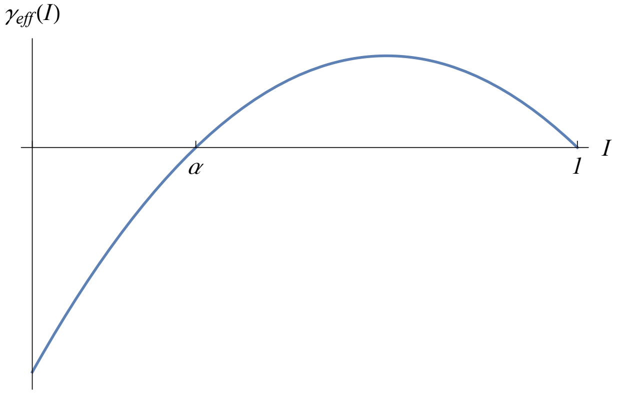


Figure 3.9: Effective local linear growth, γ_{eff}, as a function of local turbulence intensity.

in order to grow into a traveling wave front and spread; otherwise it will collapse to $I = 0$, like $e^{-|\gamma_1|t}$ (see Fig. 3.8). (If $\alpha > \alpha^*$, there is a reverse situation where a sufficiently large negative perturbation from $I = 1$ can "grow" into a wave of receding turbulence.) Numerical solutions of Eq. (3.20) suggest that there is no other possibility for the asymptotic behavior. There exists a family of nontrivial standing wave solutions, but these are linearly unstable to perturbations, as can be shown by standard arguments — see Appendix E.

It remains to estimate the threshold. The amplitude threshold is clear: The perturbation must exceed $I = \alpha$ at some point or else the effective local linear growth $\gamma_{\text{eff}} = (I - \alpha)(1 - I)$ will be negative everywhere (see Fig. 3.9).

A quantitative study of the size threshold does not appear to have been made previously. To estimate it, assume for simplicity that the initial condition is smooth and even, and that it has a single associated lengthscale (say the FWHM) L and a single maximum I_0 at $x = 0$. Consider the "cap" of the initial condition, i.e. the part exceeding $I = \alpha$, as shown in Fig. 3.10. The threshold is determined by a competition between the outgoing diffusive flux from the cap and the total nonlinear growth of the turbulent intensity within the cap. The turbulent mass diffused out of the cap will enter a region where γ_{eff} is negative and will be dissipated. This competition is suggested by the form of the Lyapunov functional, Eq. (3.22) — indeed, if the potential term is larger than the kinetic term and $\mathscr{F}[I] < 0$, the initial data must grow due to the monotonicity of $\mathscr{F}$, since

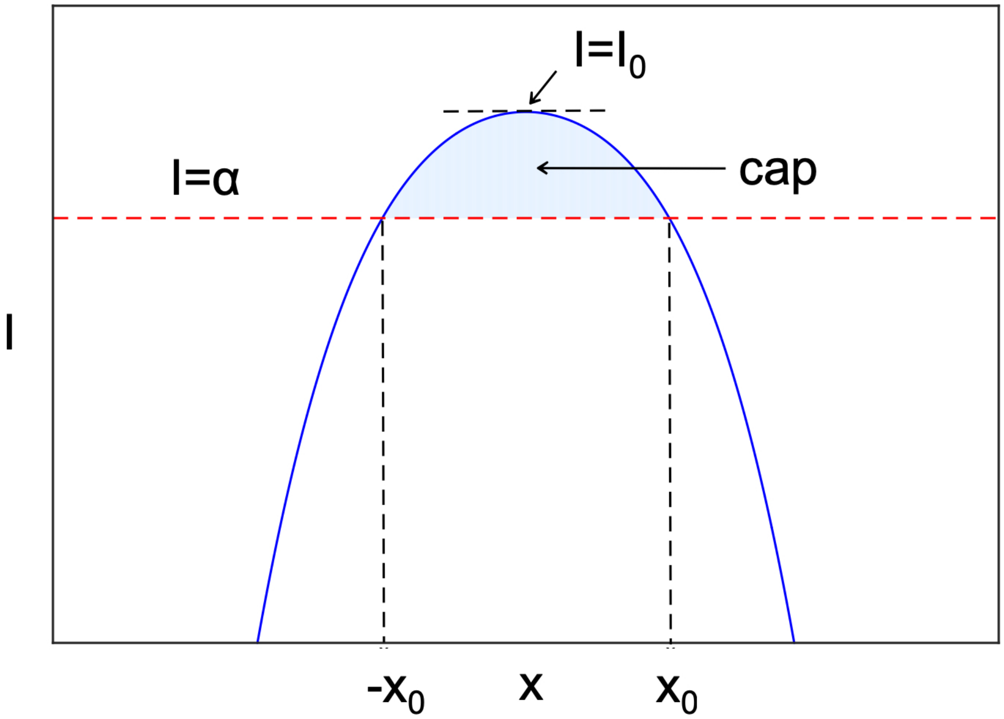


Figure 3.10: The "cap" of an initial condition. The competition between nonlinear turbulence growth in the cap and diffusive flux out of the cap at $x = \pm x_0$ generates a threshold lengthscale for growth of the initial condition. Reproduced from Heinonen and Diamond, *Physics of Plasmas* **26**, 030701 (2019), with the permission of AIP Publishing.

$\mathscr{F}[I=0] = 0$ and $\mathscr{F}[I=1] = -\infty$. The competition sets a minimum lengthscale for spreading $\sqrt{D/\gamma}$. Thus, diffusion both drives the "avalanche" by providing the spreading mechanism and inhibits it by increasing the threshold size of the slug.

We now write $I''(0) = -I_0/L^2$ (assume this derivative is nonzero). This effectively defines the lengthscale L. In the cap, we have

$$I(x) = I_0 - \frac{I_0}{2L}x^2 + O(x^4) \tag{3.33}$$

and $I(x)$ crosses $I = \alpha$ at $\pm x_0$, where

$$x_0 \simeq \sqrt{2\left(1 - \frac{\alpha}{I_0}\right)} L. \tag{3.34}$$

We now integrate Eq. (3.20) from -x_0 to x_0. Expanding

$$\begin{aligned}\int_{-x_0}^{x_0} dx\, f(I) &\simeq \int_{-x_0}^{x_0} dx \left[f(I_0) - \frac{I_0}{2L^2} x^2 f'(I_0)\right] \\ &= 2x_0 f(I_0) - \frac{I_0}{3L^2} {x_0}^3 f'(I_0),\end{aligned} \tag{3.35}$$

one finds that the turbulent mass in the cap, $\mathscr{I} \equiv \int_{-x_0}^{x_0} dx I$, evolves as

$$\partial_t \mathscr{I} \simeq 2f(I_0)x_0 - \frac{I_0}{3L^2} {x_0}^3 f'(I_0) - \frac{2D(\alpha) I_0 x_0}{L^2}, \tag{3.36}$$

The mass in the cap thus grows initially if

$$\begin{aligned}L \gtrsim L_{\min} &= \sqrt{\frac{D(\alpha) I_0}{f(I_0) - \frac{1}{3}(I_0 - \alpha) f'(I_0)}} \\ &= \sqrt{\frac{3 D \alpha I_0}{\gamma (I_0 - \alpha)((1 - 2\alpha) I_0 + \alpha)}}.\end{aligned} \tag{3.37}$$

In particular, we recover power law behavior $L_{\min} \sim (I_0 - \alpha)^{-1/2}$ near the amplitude threshold, $I_0 \gtrsim \alpha$. The above estimate agrees well with numerical solution of the PDE—see Fig. 3.11.

This argument breaks down for $\alpha > \alpha^*$, in which case (empirically speaking, according to numerical solutions) compactly supported initial conditions will always collapse eventually, regardless of whether they exceed the above threshold. This is unsurprising, as in this regime it is impossible to form a traveling wave with positive speed. The assumption that initial growth of the cap guarantees asymptotic spreading, implicitly made above, is certainly no longer reasonable.

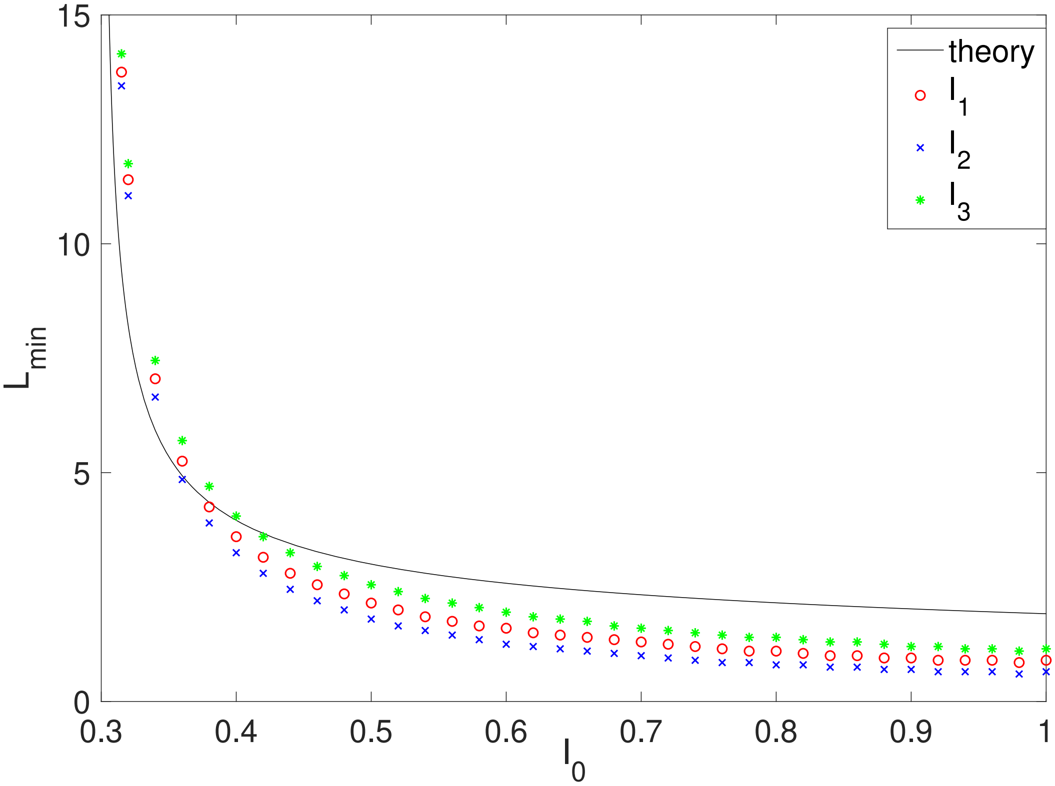


Figure 3.11: Numerically obtained lengthscale threshold $L_{\min}(I_0)$ at $\alpha = 0.3$ for three different functional forms for the initial data: $I_1 = I_0 \exp(-x^2/L^2)$ (red), $I_2 = I_0/(1+x^2/L^2)$ (blue), and $I_3 = I_0(1-\frac{x^2}{L^2})$ supported on $|x| < L$ (green). The numerical results are compared with Eq. (3.37) (solid line). $L_{\min}$ is expressed in units of $\sqrt{D/\gamma}$. Reproduced from Heinonen and Diamond, *Physics of Plasmas* **26**, 030701 (2019), with the permission of AIP Publishing.

The physical consequences of this double threshold in the MF plasma system are of interest. In the regime where $\alpha > \alpha^*$, the lengthscale threshold is effectively infinite; the system acts as if properly linearly damped and is stable to turbulent fluctuations. For smaller α, things are more interesting. In order for an initially localized slug to spread, it must have peak intensity $I > I_-$, *and* it must be correlated over a length $\ell \gtrsim L_{\min} \sim (D_0/\gamma_3 I_+)^{1/2}$.

How might this threshold be exceeded in a real system? One possibility is noise. In the weak linear damping limit $|\gamma_1| \ll \gamma_2{}^2/\gamma_3$, we have

$$I_- \sim \frac{|\gamma_1|}{\gamma_2} \tag{3.38}$$

and

$$L_{\min} \sim \left(\frac{D_0}{\gamma_2}\right)^{1/2} \sim \left(\frac{\chi_{GB}}{\omega_*}\right)^{1/2} \sim \Delta_c, \tag{3.39}$$

where Δ_c is the correlation length of the turbulence. $\frac{|\gamma_1|}{\gamma_2}$ is small in this limit, and Δ_c is the smallest lengthscale of interest, so these are meager constraints. This suggests front propagation may be triggered by noise near the marginal stability threshold. One possibility for a noise term is of the form $(I(x,t)+I_0)\eta(x,t)$, where η is Gaussian white noise. Here I_0 represents a small background, say due to sub-ion scale turbulence. The multiplicative noise $\propto I(x,t)$ is a simple, reasonable choice that (correctly) vanishes in the absence of fluctuations. Numerical solutions of the stochastic PDE (SPDE), i.e. Eqs. (3.10) and (3.11) with the noise term, with initial data $I(x,t=0)=0$, indicate that small slugs of turbulence will spontaneously form and grow as a result of the multiplicative noise. As compared to linear diffusion $\propto \partial_{xx} I$, the nonlinear diffusion is ineffective at smoothing out the slugs because the term $\propto I\partial_{xx} I$ vanishes at small amplitude. Close to marginality, slugs can exceed the threshold and form a propagating front. This is consistent with the bursty, intermittent character of avalanching. Note that the turbulence transition in pipe flow is similarly intermittent [PM80]. This behavior is illustrated in Fig. 3.12, a numerical solution to the SPDE on a box of size $L = 10$ with $\alpha = 0.05, D = 1, \gamma = 10, I_0 = 0.01$,

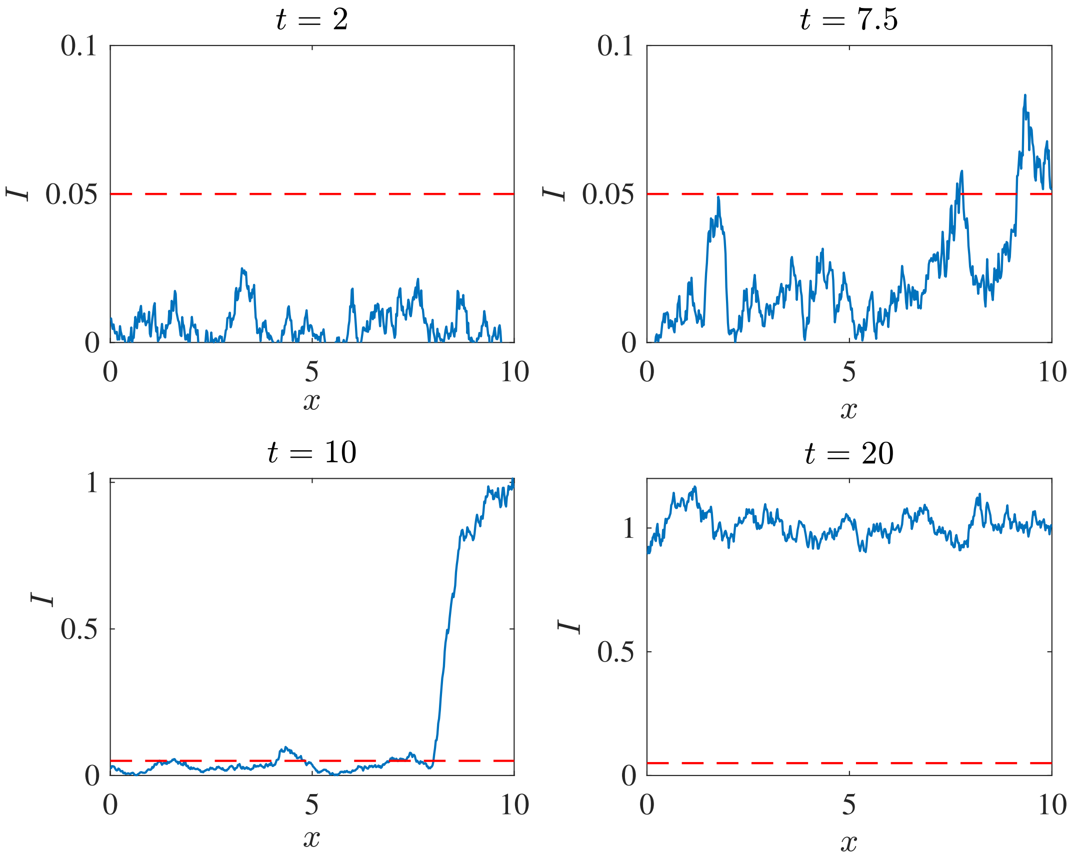


Figure 3.12: Snapshots of numerical solution of the SPDE on a box of size $L = 10$, using a semi-implicit Milstein method[LPS14]. We choose the parameters $\alpha = 0.05, D = 1, \gamma = 10, I_0 = 0.01$, and $\langle \eta^2 \rangle = 0.04$ and use a homogeneous Neumann boundary condition. Localized slugs of turbulence spontaneously grow over time until they exceed threshold (dashed red line), at which point then grow into a noisy front. Eventually, the entire system is saturated.

and $\langle \eta^2 \rangle = 0.04$.

Another possibility, specific to the $E \times B$ staircase, is that slugs can grow at temperature corrugation locations due to local exceedance of the linear stability threshold [GD17]. We suggest that if they reach a critical size corresponding to Eq. (3.37), they may cascade into the steps of the staircase via nonlinear instability, causing an avalanche. Indeed, avalanche activity between the corrugations is seen in simulations [DPHG$^+$15].

3.3.6 Analysis: penetration into a bistable zone

We now consider the spillover of turbulence into the linearly stable zone. In analogy to the Fisher case, we take

$$\begin{cases} \partial_t I = \gamma_g I + \gamma_2 I^2 - \gamma_3 I^3 + \partial_x (D_0 I \partial_x I), & x < 0 \\ \partial_t I = -\gamma_d I + \gamma_2 I^2 - \gamma_3 I^3 + \partial_x (D_0 I \partial_x I), & x > 0 \end{cases} \tag{3.40}$$

with all the γ's positive and $\gamma_d < {\gamma_2}^2 / 4\gamma_3$. This models the spreading of turbulence as it passes from a weakly supercritical region to a weakly subcritical one, in the presence of a bistability-inducing effect (such as temperature corrugations). It can also be seen as modeling avalanche penetration into a subcritical region.

For initial conditions supported on $x < 0$, a traveling wave with turbulence amplitude $A = \lim_{x \to -\infty} I(x - ct) = (\gamma_2 + \sqrt{{\gamma_2}^2 + 4\gamma_g \gamma_3})/2\gamma_3$ will develop in the unstable zone, which will eventually penetrate the linearly stable zone $x > 0$.

Generally speaking, when $\alpha < \alpha^*$ the bistable zone will behave roughly like a linearly *unstable* Fisher system, and when $\alpha > \alpha^*$ it will behave roughly like a linearly *stable* Fisher system, consistent with the picture developed in the preceding sections.

In particular, since the wave amplitude A must be greater than the value of the unstable root I_- in the stable zone, one expects for $\gamma_d < \frac{15{\gamma_2}^2}{64\gamma_3}$ (i.e. $\alpha < \alpha^*$, which is most of the acceptable range for γ_d) that a second traveling wave with reduced amplitude $A' = (\gamma_2 + \sqrt{\gamma_2^2 - 4\gamma_d \gamma_3})/2\gamma_3$ and reduced speed approximated by Eq. (3.32).

For $\frac{15}{64} < \frac{\gamma_d \gamma_3}{{\gamma_2}^2} < \frac{1}{4}$ $(\alpha > \alpha^*)$, the wave penetrates the second region only a finite depth and forms a stationary intensity profile. By direct integration and invocation of continuity of I and $\partial_x I$ at $x = 0$, an exact expression for the penetration depth λ may be obtained as an elliptic integral, but we omit this calculation since the result is fairly complicated, and the relevant parameter region is very small.

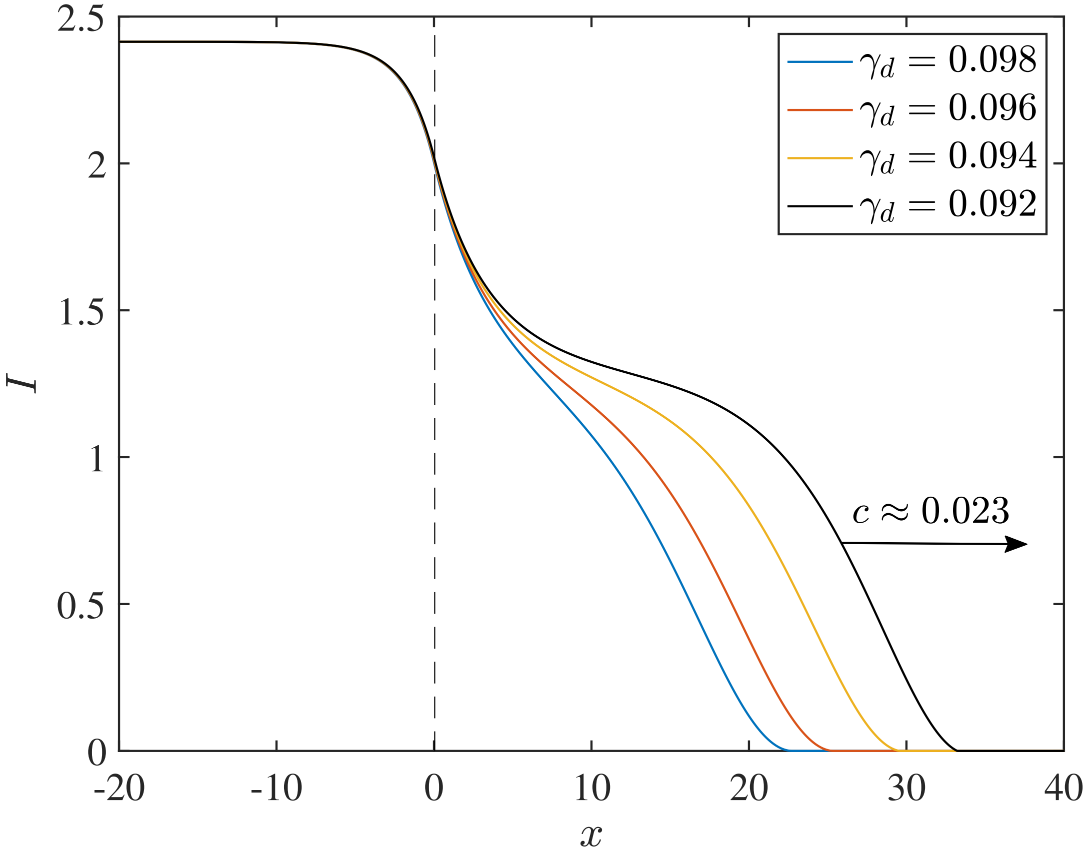


Figure 3.13: Plot showing the penetration of a traveling wave into the bistable region, for $\gamma_g = 0.1, \gamma_2 = 0.2, \gamma_3 = 0.1, D_0 = 1$ and several values of the linear damping γ_d in the second region $x > 0$. When the damping exceeds the critical value 0.09375, the wave penetrates a finite depth and forms a steady intensity profile as in the Fisher model, whereas below this value, a new traveling wave is formed in the second region.

The results of numerical solutions agree with these expectations. In Fig. 3.13, the turbulence profiles for $\gamma_g = 0.1, \gamma_2 = 0.2, \gamma_3 = 0.1$, and $D_0 = 1$ are displayed. The linearly damped region is bistable for $\gamma_d < 0.1$. As γ_d approaches the critical value $\frac{15\gamma_2{}^2}{64\gamma_3} = 0.09375$ from above, the finite penetration depth into the linearly damped region increases. Below critical γ_d, a traveling wave instead forms in the linearly damped region, with speed consistent with Eq. (3.32) (see Fig. 3.14).

We conclude that the inclusion of a bistability-inducing effect significantly enhances the spreading of turbulence into a linearly stable zone—rather than penetrating just a few correlation

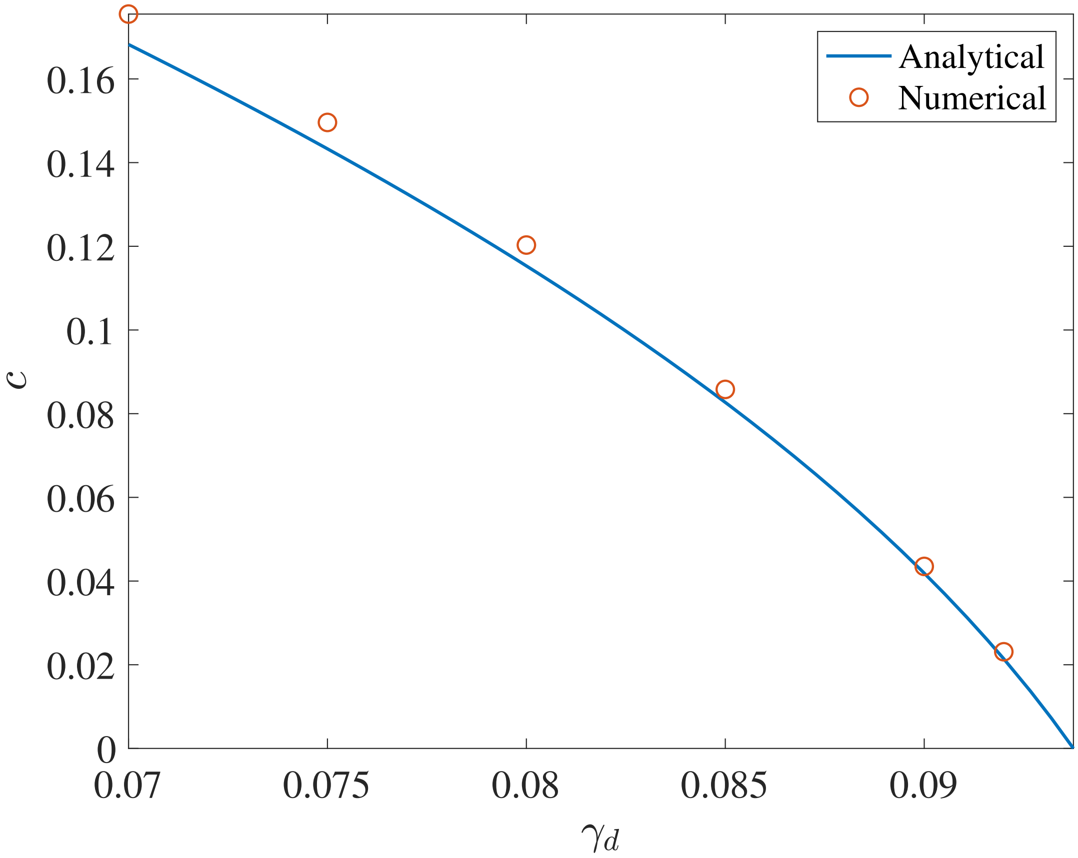


Figure 3.14: Numerically obtained speed of traveling wave in second region, for $\gamma_d < 0.09375$ and using the same conditions as in Fig. (3.13). The result shows good agreement with Mornev-Pedersen approximation [Ped05].

lengths, turbulence propagates ballistically into the linearly stable region as long as the system remains near marginal. This strongly delocalizes the flux-gradient relation. In particular, if the second region is marginally subcritical (small γ_d), the propagation speed into the stable zone is of the order $c \sim \sqrt{D_0 \tilde{\gamma}}$ (as before), independent of γ_d. It should be noted that the bistability is a result of global profile structure. The local structure alone, then, is not enough to determine how far the turbulence spreads.

3.4 Discussion and conclusion

We have introduced and analyzed a new, simple model for turbulence spreading. This model is similar to, but improves upon, the more familiar Fisher equation-based model in the ways listed below. Specifically, the model we present:

1. correctly accounts for observations of subcritical turbulence, which are incompatible with the supercritical Fisher model,
2. predicts ballistic penetration of turbulence into a linearly stable zone with weak damping, resolving the issue that Fisher fronts penetrate feebly

and

3. resolves the issue that spreading in a Fisher model is only possible in a linearly unstable zone, which is dubiously physical in the presence of noise.

This model can be seen as phenomenological, but it has a firm grounding in effects due to inhomogeneous mixing and temperature corrugations, or (more generally) in observations of subcritical turbulence.

Additionally, we have also:

4. shown that bistability admits spontaneous intermittent propagation of turbulence, providing a dynamical mechanism for avalanching,

5. computed a quantitative estimate for the threshold size of a turbulence pulse required to trigger such an avalanche

and

6. forged a direct connection between turbulence spreading and avalanching in sandpiles via the Gil and Sornette model [GS96], which is closely related to ours.

These results, we believe, together a make substantial inroad into answering the question: "What is an avalanche?"

The estimate for the slug threshold should be testable in simulation, gyrokinetic or otherwise, for example, by initializing patches of turbulence of varying sizes in a stable zone or by analyzing simulations of the $E \times B$ staircase. Also of interest for future work would be a search, via careful numerical studies, for ballistic spreading of turbulence into a weakly damped stable zone, perhaps in a similar setup to [YKDH14]. If found, this would indirectly confirm, or at least support, the presence of turbulence bistability. However, care must be taken to disambiguate this effect, due to nonlinear coupling, from spreading due to gradient propagation.

This work provides a possible lens for understanding the $E \times B$ staircase: The staircase, formed by inhomogeneous turbulent mixing, induces global bistability in the turbulence. This bistability results in a local threshold for turbulent slugs, which is expected to be intermittently exceeded, producing the avalanche activity observed in steps of the staircase.

Our model also makes a prediction of hysteresis due to the global threshold behavior. We have computed thresholds for external fluxes and hysteresis strengths. As discussed above, we strongly suggest more experiments similar to Inagaki *et al.* (2013) be conducted to test this important prediction.

As already noted, this model is similar to models for the transition to turbulence in pipe flow. Pomeau [Pom86, Pom15] argued that the transition to fluid turbulence requires subcritical bifurcation in order to allow for the coexistence of turbulent and laminar domains.

He suggested a bistable reaction-diffusion equation with noise as a simple model for the local fluctuation amplitude. It is not surprising, then, that the transition to fluid turbulence is marked by intermittent bursts [PM80]. These can be seen as avalanches, triggered by the spontaneous growth of turbulent domains to a critical size.

Finally, it is important to acknowledge that a complete model for turbulence spreading must include coupling to zonal flows and address the effects of spreading on profiles. Zonal flows are known to interact with and regulate turbulence in a predator–prey dynamic [DIIH05b], and it is reasonable to expect that they would have a quenching effect that allows the system to relax after being excited. Recent work [SD20] has addressed the question of the effect of turbulence spreading on profiles. Results indicate that spreading has a significant impact when the drive and profile structure support a sustained and sharp intensity gradient. One clear case of this is in H-mode in the turbulent layer immediately adjacent to the pedestal. Spreading allows a more robust saturation process, which actually yields an *increased* pedestal height. This translates into an overall improvement in confinement. Future research will focus on incorporating the effects of zonal flow, as well as further studies of profile evolution.

Chapter 3, in full, is a reprint of the material as it appears in R. A. Heinonen and P. H. Diamond, *Physics of Plasmas* 27, 032303 (2020). The dissertation author was the primary investigator and author of this paper.

Chapter 4

On the role of cross-helicity in β-plane magnetohydrodynamic turbulence

4.1 Introduction

Since the early 1990s, much attention has been paid to the *solar tachocline*, a thin layer which lies at the base of the convective zone (CZ) of the sun. The tachocline separates the rigidly rotating core from the differentially rotating CZ and exhibits strong radial shear. It is widely believed [Cha20] that the tachocline is home to the so-called Ω-effect and thus of crucial importance to the solar dynamo: the shear drags poloidal magnetic field lines originating from the core, converting them to a strong toroidal field, which is stored in the tachocline.

Considerable debate in the literature has surrounded the nature of turbulent momentum transport in the tachocline, which is of central importance to the question of why the tachocline exists in the first place. Spiegel and Zahn (1992) [SZ92] first proposed that the turbulence would act as an eddy viscosity, which would prevent the tachocline from spreading inward due to radiative diffusion. Gough and McIntyre (1998) [GM98] instead argued that the turbulence would act as a *negative* viscosity and produce mean flows through potential vorticity mixing, which opposes the downward "burrowing" of gas from the convection zone by meridional circulation.

It is thus fruitful to study meso- and macro-scale turbulent momentum transport in the tachocline. Because the lower tachocline is stably stratified, it can be modeled as a quasi-two-dimensional, differentially rotating magnetic fluid, which motivates the consideration of a magnetohydrodynamic (MHD) model on a 2-D β-plane:

$$\partial_t \nabla^2 \psi + \beta \partial_x \psi = \{\psi, \nabla^2 \psi\} - \{A, \nabla^2 A\} + \nu \nabla^4 \psi + \tilde{f} \tag{4.1}$$

$$\partial_t A = \{\psi, A\} + \eta \nabla^2 A + \tilde{g}. \tag{4.2}$$

These incompressible equations express the dynamics of the streamfunction ψ and the magnetic potential A, defined so that the velocity field is $\mathbf{v} = (\partial_y \psi, -\partial_x \psi, 0)$ and the magnetic field is $\mathbf{b} = (\partial_y A, -\partial_x A, 0)$. We have defined the Poisson bracket $\{a, b\} = \partial_x a \partial_y b - \partial_y a \partial_x b$. β is the so-called Rossby parameter, which captures the effect of a planetary vorticity/Coriolis force gradient, so that the solar rotation rate is locally given by $2\Omega \simeq (0, 0, f_0 + \beta y)$. ν is the fluid viscosity and η is the magnetic diffusivity. At the moment we do not specify the forcing functions $\tilde{f}$ and $\tilde{g}$, which respectively model kinetic stirring (including thermal excitation) and injection of magnetic potential, say by pumping from above or below the tachocline. It is worth noting that we have normalized the magnetic field by $1/\sqrt{4\pi\rho}$ so that it has units of velocity.

This model, introduced by Tobias *et al.* (2007) [TDH07] and Diamond *et al.* (2007) [DIIS07], constitutes a major simplification of the physics of the tachocline, but it nevertheless features a quite rich range of dynamics and is fundamentally interesting in its own right. Tobias *et al.* studied the effect of a weak mean magnetic field aligned with the shear, so that $A \to \tilde{A} + b_0 y$, and showed numerically that the system undergoes a transition from Alfvénic turbulence with a forward cascade to Rossby-like turbulence featuring an inverse cascade to large scales and zonal flows, when the magnetic diffusivity becomes sufficiently large (or the mean field becomes sufficiently weak). Diamond *et al.* (2007) supplemented this study with an analytic closure calculation and derived the “magnetic Rhines scale” $\ell_{\mathrm{MR}} = \sqrt{b_0/\beta}$, defined by the crossover of the Rossby frequency $\omega_\beta = -\beta k_x/k^2$ and the Alfvén frequency $\omega_A = k_x b_0$. They argued that ℓ_{MR} is a critical lengthscale below which the turbulence is Alfvénized, so that the Maxwell-Reynolds stress

$$\langle \partial_x \tilde{\psi} \partial_y \tilde{\psi} \rangle - \langle \partial_x \tilde{A} \partial_y \tilde{A} \rangle \simeq \int d^2\mathbf{k} \, \frac{k_x k_y}{k^2} (|\tilde{\mathbf{v}}_{\mathbf{k}}|^2 - |\tilde{\mathbf{b}}_{\mathbf{k}}|^2) \tag{4.3}$$

vanishes to leading order, and there is no clear large-scale momentum transport to set up a zonal flow. Conversely, at scales larger than the magnetic Rhines scale, the turbulence is Rossby-like. Thus, it is natural to suspect that the system will undergo transition from Alfvénic to Rossby turbulence when the magnetic Rhines scale becomes smaller than typical scales of the problem.

However, omitted from previous studies is a reckoning with the fact that the Coriolis term explicitly breaks conservation of cross-helicity, that is the alignment of the magnetic field and velocity $H = \langle \tilde{\mathbf{v}} \cdot \tilde{\mathbf{b}} \rangle = -\langle \tilde{A} \nabla^2 \tilde{\psi} \rangle$. Cross-helicity is globally conserved in pure MHD turbulence (up to dissipation), but when $\beta \neq 0$ we have

$$\partial_t \langle \tilde{A} \nabla^2 \tilde{\psi} \rangle = -\beta \langle \tilde{v}_y \tilde{A} \rangle + \text{dissipation}. \tag{4.4}$$

The origin of the breaking of cross-helicity conservation is that the β term, which (unlike a mean magnetic field term) cannot be removed by a simple change of variables, breaks symmetry in the $\hat{x}$ direction.

We can anticipate that a nonzero cross-helicity in this system may have important consequences for transport. Cross-helicity is closely related to the turbulent emf [Mof78], so that one may naturally suspect that the cross-helicity induced by differential rotation in the tachocline has consequences for the solar dynamo. However, it is well known that a dynamo cannot be supported in two-dimensions, so our primary subject of study will be on momentum transport, which will generally couple to the cross-helicity in this system. In fact, we will see that the relationship between cross-helicity and momentum transport in this system is quite intimate.

The link of cross helicity to transport may be expressed in terms of the spectrum $H_{\mathbf{k}} = \langle \tilde{\mathbf{v}}_{\mathbf{k}} \cdot \tilde{\mathbf{b}}_{-\mathbf{k}} \rangle$. We then have

$$\frac{k_x k_y}{2k^2} \operatorname{Re} H_{\mathbf{k}} = \hat{z} \cdot \langle \tilde{\mathbf{v}}_{\mathbf{k}} \times \tilde{\mathbf{b}}_{-\mathbf{k}} \rangle \tag{4.5}$$

$$\frac{k_x}{k^2} \operatorname{Im} H_{\mathbf{k}} = \langle \tilde{v}_{y,\mathbf{k}} \tilde{A}_{-\mathbf{k}} \rangle. \tag{4.6}$$

Thus, the real part of the cross helicity determines the turbulent emf, and the imaginary part determines the turbulent flux of magnetic potential along the parallel axis. The turbulent emf is responsible for mean field evolution.

The cross-helicity may be viewed as a dual of the alignment of Elsässer populations. If we define the Elsässer modes $\mathbf{z}^{\pm} = \tilde{\mathbf{v}} \pm \tilde{\mathbf{b}}$, the utility of this basis lies in the well-known fact that pure MHD turbulence is entirely generated by counter-propagating Elsässer populations; $\mathbf{z}^{+}$ and $\mathbf{z}^{-}$ interact with each other but not themselves. In this basis, the Maxwell-Reynolds stress, which gives rise to momentum transport, is equivalent to the *alignment* of counter-propagating Elsässers:

$$\operatorname{Re} \langle \mathbf{z}^{+}_{\mathbf{k}} \cdot \mathbf{z}^{-}_{-\mathbf{k}} \rangle = |\tilde{v}_{\mathbf{k}}|^2 - |\tilde{b}_{\mathbf{k}}|^2. \tag{4.7}$$

Similarly, cross-helicity is equivalent to an imbalance in Elsässer populations; if we define $H_{\mathbf{k}} = \langle \tilde{\mathbf{v}}_{\mathbf{k}} \cdot \tilde{\mathbf{b}}_{-\mathbf{k}} \rangle$, we have

$$\operatorname{Re} H_{\mathbf{k}} = \frac{1}{4} (|\mathbf{z}^{+}_{\mathbf{k}}|^2 - |\mathbf{z}^{-}_{\mathbf{k}}|^2) \tag{4.8}$$

In general, neither basis is preferred when the Rossby parameter is turned on. Instead, the eigenmodes are Rossby-Alfvén modes, whose dispersion combines those of simple Rossby waves and Alfvén waves:

$$\omega = \frac{\omega_\beta \pm \sqrt{4\omega_A^2 + \omega_\beta^2}}{2}, \tag{4.9}$$

where ω_β and ω_A are respectively the Rossby and Alfvén frequencies. The coupling of Rossby-Alfvén waves of different species will mix the Elsässer alignment and the cross-helicity, so that cross-helicity will play a role in all the spectral equations and help determine momentum transport.

In this work, we study the role of cross-helicity in β-plane MHD turbulence. First, we show by a simple non-perturbative calculation—arguing on the basis of conservation of squared magnetic potential $\langle \tilde{A}^2 \rangle$—that the total cross-helicity in this system attains a finite, stationary value. Second, we discuss a basic closure of the system within the framework of weak turbulence, and argue that for strong mean field, the stationary real cross-helicity spectrum is in fact equivalent to the Elsässer alignment spectrum. This argument also explicitly shows that the turbulence is magnetic below the magnetic Rhines scale and kinetic above, as argued in [DIIS07]. We also discuss the difficulties of making analytical progress from the weak turbulence spectral equations; in particular, we point out that the $\beta \to 0$ limit is singular, and thus the tempting route of expanding the weak turbulence closure around small β fails.

Next, we perform a set of simulations of this system for strong mean field and several values of β and demonstrate that

- the simple non-perturbative calculation provides an accurate estimate of the actual stationary cross-helicity,
- transition from Alfvénic to Rossby turbulence begins when the magnetic Rhines wavenumber exceeds the forcing wavenumber: $k_{\mathrm{MR}} = \sqrt{\beta / b_0} \gtrsim k_f^2$,
- and this transition is presaged by an increase in the global cross-helicity, which peaks around the critical β where the above scales overlap.

We also present plots of the stationary spectra from simulation, which confirm the Rossby-like nature of the turbulence at large β and the MHD-like nature at small β, and verifies the relationship between the cross-helicity and the Elsässer alignment within weak turbulence.

4.2 Theory

4.2.1 Stationary cross-helicity

Our investigation begins with a simple, non-perturbative calculation of the cross-helicity at long times. The calculation is inspired by the Zel'dovich theorem [Zel57]; like that earlier result, here we use the conservation of mean-squared magnetic potential to express large-scale transport properties at long times in terms of small-scale dissipation.

For the reader's convenience, we first rewrite the system (4.1)-(4.2) with the mean magnetic field terms. We assume the system is only kinetically forced ($\tilde{g} = 0$):

$$\partial_t \nabla^2 \tilde{\psi} + \beta \partial_x \tilde{\psi} = \{\tilde{\psi}, \nabla^2 \tilde{\psi}\} - \{\tilde{A}, \nabla^2 \tilde{A}\} + b_0 \partial_x \nabla^2 \tilde{A} + \nu \nabla^4 \tilde{\phi} + \tilde{f} \tag{4.10}$$

$$\partial_t \tilde{A} = \{\tilde{\psi}, \tilde{A}\} + b_0 \partial_x \tilde{\psi} + \eta \nabla^2 \tilde{A} \tag{4.11}$$

Multiplying the second equation by A, integrating over space, and taking $t \to \infty$ yields

$$\frac{1}{2} \partial_t \langle \tilde{A}^2 \rangle = b_0 \langle \tilde{A} \partial_x \tilde{\psi} \rangle - \eta \langle (\nabla \tilde{A})^2 \rangle = 0 \tag{4.12}$$

so the mean flux of magnetic potential is

$$\langle \tilde{A} \partial_x \tilde{\psi} \rangle = \frac{\eta}{b_0} \langle \tilde{b}^2 \rangle. \tag{4.13}$$

This is just Zel'dovich's theorem, which also requires that the *total* mean magnetic field must eventually decay in 2D. However, this decay may be very slow [Pou78].

Now, multiplying the A equation by $\nabla^2 \psi$ and vice-versa, integrating over space, and summing the resulting equations yields

$$\partial_t \langle \tilde{A} \nabla^2 \tilde{\psi} \rangle = -\beta \langle \tilde{A} \partial_x \tilde{\psi} \rangle + (\eta + \nu) \langle \nabla^2 \tilde{\psi} \nabla^2 \tilde{A} \rangle + \langle \tilde{A} \tilde{f} \rangle. \tag{4.14}$$

We then introduce characteristic lengthscales of the velocity and magnetic field variation and $\langle \nabla^2 \psi \nabla^2 A \rangle \simeq \frac{1}{\ell_v \ell_b} \langle \mathbf{v} \cdot \mathbf{b} \rangle$. Finally, we substitute (4.13) into the above, take the long-time limit, and obtain the estimate for the stationary cross-helicity

$$H(t=\infty) = \frac{\ell_v \ell_b}{\eta + \nu} \left(\frac{\beta \eta}{b_0} \langle \tilde{b}^2 \rangle - \langle \tilde{A} \tilde{f} \rangle \right). \tag{4.15}$$

Then, assuming that $\langle \tilde{A} \tilde{f} \rangle$ is small, we have the simpler expression

$$H(t=\infty) = \frac{\ell_v \ell_b}{1 + \mathrm{Pm}} \frac{\beta \langle \tilde{b}^2 \rangle}{b_0}. \tag{4.16}$$

Here, $\mathrm{Pm} \equiv \nu / \eta$ is the magnetic Prandtl number. In practice, the forcing term $\langle A \tilde{f} \rangle$ will make a nonzero contribution to the total cross-helicity, which may be significant if the magnetic Reynolds number is very large. However, in the simulations performed in this work, this contribution was relatively small, $\lesssim 15\%$ of that contributed by the β term. If the system is magnetically forced, there will be additional contributions from correlations with $\tilde{g}$.

When β is not too strong, $\beta \lesssim b_0 k_f^2$, we can estimate $\ell_v = \ell_b \simeq k_f^{-1}$. When β is large, this is no longer a good estimate for ℓ_b, as the magnetic field goes to small scales; a better estimate is then a magnetic Taylor microscale $\sqrt{\frac{\eta}{\varepsilon} \langle \tilde{b}^2 \rangle}$, where ε is the energy injection rate (equivalent to the dissipation rate for stationary turbulence).

Note that this calculation depended on the magnetic field being latitudinally aligned; otherwise, the Ze'ldovich theorem is not enough to discern the flux of magnetic potential across the planetary vorticity gradient. Note also that the cross-helicity could be very large in the case of a weak mean field and strong fluctuations.

4.2.2 Weak turbulence closure of spectral equations

We have derived an estimate for the stationary *mean* cross-helicity; however, discerning transport properties requires studying the spectra. To this end, we seek a statistical closure which treats cross-correlations like the cross-helicity spectrum on an equal footing with autocorrelations like $\langle|\tilde{v}_{\mathbf{k}}|^2\rangle$. The simplest approach, which we adopt here, is the weak turbulence theory of Sagdeev and Galeev [SG69, ZLF12], though we note that [GLP83] took an approach (in pure MHD) based on the eddy-damped quasi-normal Markovian (EDQNM) closure theory. Formally a second-order time-dependent perturbation theory, weak turbulence assumes that the linear timescales are fast compared to nonlinear ones, so that the nonlinearity may be decomposed in terms of resonant triplet interactions between linear eigenmodes — here, Rossby-Alfvén modes. For this approach to be valid, we must assume a strong mean magnetic field; even here, the weak turbulence description will be dubious at small parallel lengthscales, where the linear frequencies vanish.

In $\mathbf{k}$-space, the β-plane MHD system is expressed as

$$\partial_t \psi_{\mathbf{k}} + i\omega_{\beta,\mathbf{k}}\psi_{\mathbf{k}} + i\omega_{A,\mathbf{k}}A_{\mathbf{k}} = \frac{1}{2}\int d^2\mathbf{k}' d^2\mathbf{k}''\, \delta(\mathbf{k}-\mathbf{k}'-\mathbf{k}'')\left[(\mathbf{k}'\times\mathbf{k}'')\cdot\hat{z}\right]\frac{k'^2-k''^2}{k^2}\left(\psi_{\mathbf{k}'}\psi_{\mathbf{k}''} - A_{\mathbf{k}'}A_{\mathbf{k}''}\right)$$

$$\partial_t A_{\mathbf{k}} + i\omega_{A,\mathbf{k}}\psi_{\mathbf{k}} = \frac{1}{2}\int d^2\mathbf{k}' d^2\mathbf{k}''\, \delta(\mathbf{k}-\mathbf{k}'-\mathbf{k}'')\left[(\mathbf{k}'\times\mathbf{k}'')\cdot\hat{z}\right]\left(A_{\mathbf{k}'}\psi_{\mathbf{k}''} - \psi_{\mathbf{k}'}A_{\mathbf{k}''}\right),$$

where $\omega_{\beta,\mathbf{k}} = -\beta k_x/k^2$ is the Rossby frequency and $\omega_{A,\mathbf{k}} = b_0 k_x$ is the Alfvén frequency, and we have neglected forcing and dissipation terms.

The linear eigenmodes of this system are the two Rossby-Alfvén modes with frequencies

$$\omega^{\pm} = \frac{\omega_\beta \pm \Omega}{2}$$

and corresponding amplitudes

$$\phi^{\pm} = \frac{1}{\Omega}(\omega^{\pm}\psi - \omega_A A), \tag{4.17}$$

where $\Omega = \mathrm{sgn}(k_x)\sqrt{4\omega_A^2 + \omega_\beta^2}$. The sign of the square root and the normalization are chosen so that the $k_x \to 0$ limit is not problematic. In the $b_0 \to \infty$ limit, these are just Elsässer modes; as $b_0 \to 0$, one mode becomes a Rossby wave and the other ceases to exist (the amplitude and frequency simultanenously vanish).

The classical weak turbulence theory, formulated for a single scalar field, is exceedingly well-known [SG69, ZLF12]. However, its straightforward generalization to the case of multiple interacting scalar fields is rarely seen in the literature, so we present it here. Let $\{\phi^\alpha\}$ be a finite set of quadratically interacting linear eigenmodes described by the evolution equations

$$\partial_t \phi_{\mathbf{k}}^\alpha + i\omega_{\mathbf{k}}^\alpha \phi_{\mathbf{k}}^\alpha = \sum_{\beta\gamma} \frac{1}{2} \int d^2\mathbf{k}' d^2\mathbf{k}''\, \delta(\mathbf{k} - \mathbf{k}' - \mathbf{k}'') M_{\mathbf{k},\mathbf{k}',\mathbf{k}''}^{\alpha\beta\gamma} \phi_{\mathbf{k}'}^\beta \phi_{\mathbf{k}''}^\gamma, \tag{4.18}$$

so that the linear frequency matrix $\omega^{\alpha\beta}$ is diagonal. The coupling coefficients are assumed to obey the symmetry condition $M_{\mathbf{k},\mathbf{k}',\mathbf{k}''}^{\alpha\beta\gamma} = M_{\mathbf{k},\mathbf{k}'',\mathbf{k}'}^{\alpha\gamma\beta}$.

We define the correlators $C_{\mathbf{k}}^{\alpha\alpha'}$ by

$$\langle \phi_{\mathbf{k}}^\alpha \phi_{\mathbf{k}'}^{\alpha'} \rangle \equiv C_{\mathbf{k}}^{\alpha\alpha'} \delta(\mathbf{k} + \mathbf{k}') e^{-i(\omega_{\mathbf{k}}^\alpha - \omega_{\mathbf{k}}^{\alpha'})t}.$$

After some work, it can be shown (see Appendix F) that within the standard assumptions of weak turbulence (second-order perturbation theory, invoking both homogeneity and the random phase approximation), the correlators evolve according to the spectral equations

$$\begin{aligned}\partial_t C_{\mathbf{k}}^{\alpha\alpha'} = \sum_{\beta\gamma} \int d^2\mathbf{k}' d^2\mathbf{k}''\, \delta(\mathbf{k} - \mathbf{k}' - \mathbf{k}'') \Bigg[&\pi |M_{\mathbf{k},\mathbf{k}',\mathbf{k}''}^{\alpha\beta\gamma}|^2 C_{\mathbf{k}'}^{\beta\beta} C_{\mathbf{k}''}^{\gamma\gamma} \delta(\omega_{\mathbf{k}}^\alpha - \omega_{\mathbf{k}'}^\beta - \omega_{\mathbf{k}''}^\gamma) \delta_{\alpha\alpha'} \\ &+ M_{\mathbf{k},\mathbf{k}',\mathbf{k}''}^{\alpha\beta\gamma} M_{\mathbf{k}',\mathbf{k},-\mathbf{k}''}^{\beta\alpha\gamma} C_{\mathbf{k}}^{\alpha\alpha'} C_{\mathbf{k}''}^{\gamma\gamma} \left(\pi\delta(\omega_{\mathbf{k}}^\alpha - \omega_{\mathbf{k}'}^\beta - \omega_{\mathbf{k}''}^\gamma) + i\mathscr{P} \frac{1}{\omega_{\mathbf{k}}^\alpha - \omega_{\mathbf{k}'}^\beta - \omega_{\mathbf{k}''}^\gamma} \right) \\ &+ M_{\mathbf{k},\mathbf{k}',\mathbf{k}''}^{\alpha'\beta\gamma *} M_{\mathbf{k}',\mathbf{k},-\mathbf{k}''}^{\beta\alpha'\gamma *} C_{\mathbf{k}}^{\alpha\alpha'} C_{\mathbf{k}''}^{\gamma\gamma} \left(\pi\delta(\omega_{\mathbf{k}}^{\alpha'} - \omega_{\mathbf{k}'}^\beta - \omega_{\mathbf{k}''}^\gamma) - i\mathscr{P} \frac{1}{\omega_{\mathbf{k}}^{\alpha'} - \omega_{\mathbf{k}'}^\beta - \omega_{\mathbf{k}''}^\gamma} \right) \Bigg].\end{aligned} \tag{4.19}$$

Here, $\mathscr{P}$ means the Cauchy principal value is taken, and $\delta_{\alpha\alpha'}$ is a Kronecker delta. Note that in the case of a single mode with real coupling coefficients, the principal value terms cancel and we obtain the usual Sagdeev-Galeev theory. The Dirac deltas constrain the collision integral to resonant surfaces where the mismatch frequency $\Delta\omega = \omega^{\alpha}_{\mathbf{k}} - \omega^{\beta}_{\mathbf{k}'} - \omega^{\gamma}_{\mathbf{k}''}$ vanishes.

To apply this result to our β-plane MHD system, where we have two scalar fields ψ and A, we will will need to transform to the eigenbasis — the two Rossby-Alfvén modes, $+$ and $-$ — and obtain the coupling coefficients in this basis. These are

$$M^{\pm++}_{\mathbf{k},\mathbf{k}',\mathbf{k}''} = \frac{\hat{z}\cdot(\mathbf{k}'\times\mathbf{k}'')}{\Omega_{\mathbf{k}}}\left[\frac{k_x\omega^{-}_{\mathbf{k}'}}{k'_x} - \frac{k_x\omega^{-}_{\mathbf{k}''}}{k''_x} + \frac{k''^2-k'^2}{k^2}\omega^{\pm}_{\mathbf{k}}\left(1-\frac{\omega^{-}_{\mathbf{k}'}\omega^{-}_{\mathbf{k}''}}{k'_x k''_x b_0^2}\right)\right] \tag{4.20}$$

$$M^{\pm+-}_{\mathbf{k},\mathbf{k}',\mathbf{k}''} = -\frac{\hat{z}\cdot(\mathbf{k}'\times\mathbf{k}'')}{\Omega_{\mathbf{k}}}\left[\frac{k_x\omega^{-}_{\mathbf{k}'}}{k'_x} - \frac{k_x\omega^{+}_{\mathbf{k}''}}{k''_x} + \frac{k''^2-k'^2}{k^2}\omega^{\pm}_{\mathbf{k}}\left(1-\frac{\omega^{-}_{\mathbf{k}'}\omega^{+}_{\mathbf{k}''}}{k'_x k''_x b_0^2}\right)\right] \tag{4.21}$$

$$M^{\pm--}_{\mathbf{k},\mathbf{k}',\mathbf{k}''} = \frac{\hat{z}\cdot(\mathbf{k}'\times\mathbf{k}'')}{\Omega_{\mathbf{k}}}\left[\frac{k_x\omega^{+}_{\mathbf{k}'}}{k'_x} - \frac{k_x\omega^{+}_{\mathbf{k}''}}{k''_x} + \frac{k''^2-k'^2}{k^2}\omega^{\pm}_{\mathbf{k}}\left(1-\frac{\omega^{+}_{\mathbf{k}'}\omega^{+}_{\mathbf{k}''}}{k'_x k''_x b_0^2}\right)\right]. \tag{4.22}$$

We also may express the Rossby-Alfvén correlators in the basis $E^K_{\mathbf{k}} = \langle|v_{\mathbf{k}}|^2\rangle, E^M_{\mathbf{k}} = \langle|b_{\mathbf{k}}|^2\rangle, H_{\mathbf{k}}:$

$$k^2 C^{\pm\pm}_{\mathbf{k}} = \frac{1}{\Omega^2}\left(\omega_{\pm}^2 E^K_{\mathbf{k}} + \omega_A^2 E^M_{\mathbf{k}} - 2\omega_A\omega_{\pm}\,\mathrm{Re}\,H_{\mathbf{k}}\right) \tag{4.23}$$

$$k^2\,\mathrm{Re}(C^{+-}_{\mathbf{k}}e^{-i\Omega t}) = -\frac{1}{\Omega^2}\left(\omega_A^2(E^K_{\mathbf{k}} - E^M_{\mathbf{k}}) + \omega_\beta\omega_A\,\mathrm{Re}\,H_{\mathbf{k}}\right) \tag{4.24}$$

$$k^2\,\mathrm{Im}(C^{+-}_{\mathbf{k}}e^{-i\Omega t}) = -\frac{\omega_A}{2\Omega}\,\mathrm{Im}\,H_{\mathbf{k}}. \tag{4.25}$$

We see that the (real) cross-correlator between Rossby-Alfvén modes of opposite sign mixes the (real) cross-helicity and the Elsässer alignment $E^K_{\mathbf{k}} - E^M_{\mathbf{k}}$, and naturally oscillates at frequency $\Omega = \omega^{+} - \omega^{-}$.

In principle, one may proceed by using the above expressions for the coupling coefficients, the transformation from the Rossby-Alfvén basis to the physical basis, and the inverse of that transformation to transform the spectral equations Eq. 4.19 into the physical basis. However, the resulting equations are exceedingly complicated and not terribly enlightening by inspection.

It is tempting, then, to consider the limit $\beta \to 0$ and expand about a pure (weak) MHD fixed point (say, one where the energy spectra are flat and the cross helicity is zero) [TNG13], in order to perturbatively solve for the stationary spectra in powers of β, in the spirit of mean-field electrodynamics. However, this approach fails. To see this, consider the Rossby-Alfvén wave-wave interactions that contribute to the collision integrals. In pure MHD, only counter-propagating Elsässer populations interact, so the only processes with nonzero amplitude are $\pm, \mp \to \pm$. The linear frequencies may be expanded as

$$\omega_{\pm} = \pm \omega_A + \frac{\omega_\beta}{2} + O(\beta^2), \tag{4.26}$$

so the mismatch frequencies, which set the interaction time in weak turbulence, for these processes are

$$\Delta\omega = \omega^{\pm}_{\mathbf{k}} - \omega^{\pm}_{\mathbf{k}'} - \omega^{\mp}_{\mathbf{k}''} \simeq \pm 2 b_0 k''_x + \frac{1}{2}(\omega_{\beta,\mathbf{k}} - \omega_{\beta,\mathbf{k}'} + \omega_{\beta,\mathbf{k}''}), \tag{4.27}$$

where we have used the momentum conservation constraint $\mathbf{k}' + \mathbf{k}'' = \mathbf{k}$. These processes present no particular issue for perturbation theory.

When β is nonzero, previously forbidden processes now have nonzero amplitude. First, there is $\pm, \pm \to \pm$, with mismatch frequency

$$\Delta\omega \simeq \frac{1}{2}(\omega_{\beta,\mathbf{k}} - \omega_{\beta,\mathbf{k}'} - \omega_{\beta,\mathbf{k}''}), \tag{4.28}$$

equivalent to an effective Rossby-wave interaction. This frequency is $O(\beta)$, so the interaction is long-lived, $\tau \sim 1/|\beta|$. These processes pick up a coefficient $M^2 \sim \beta^2$, so their overall contribution to the collision integrals is $O(\beta)$, still not necessarily problematic in perturbation theory.

However, the $\pm,\pm \to \mp$ processes have mismatch frequency

$$\Delta\omega \simeq \pm 2b_0k_x + \frac{1}{2}(\omega_{\beta,\mathbf{k}} - \omega_{\beta,\mathbf{k}'} - \omega_{\beta,\mathbf{k}''}). \tag{4.29}$$

One would like to make the formal expansion

$$\delta(\Delta\omega_0 + \beta\Delta\omega_1 + \dots) \simeq \Delta\omega_0 + \beta\Delta\omega_1\delta'(\Delta\omega_0) + \dots, \tag{4.30}$$

where δ' is the distributional derivative of the Dirac delta function, which may be handled in practice using integration by parts. This expansion is impossible for this process because the zeroth-order mismatch frequency $\Delta\omega_0 = \pm 2b_0k_x$ is independent of the integration variable $\mathbf{k}'$, and so the distributional derivative is undefined — any attempt to perform the collision integral will introduce a divergent factor $\left|\frac{\partial\Delta\omega_0}{\partial\mathbf{k}'}\right|^{-1}$.

Physically, this issue originates from a change in the topology of the resonance surfaces for the $\pm,\pm \to \mp$ processes when β is turned on. When β is zero, the resonance surface is effectively the entire plane (k'_x, k'_y), as the resonance condition may be satisfied for any $\mathbf{k}'$ as long as the resultant wave has $k_x = 0$. When β is nonzero, the resonance surface collapses to a pair of closed curves when $k_x \neq 0$ and a pair of lines when $k_x = 0$ — see Fig. 4.1.

Another interpretation of this is that all triplet Rossby-Alfvén processes *except* $\pm,\pm \to \mp$ may be decomposed into interactions of pure Alfvén and/or Rossby waves.

In any case, it is thus difficult to make direct analytic progress with the weak turbulence spectral equations. However, one realizes that in a stationary state, the time-averaged Rossby-Alfvén cross-correlator must *vanish* as long as $k_x \neq 0$ since it naturally oscillates at the fast frequency Ω. This imposes a major constraint on the stationary, time-averaged spectra: we must have

$$E^K_{\mathbf{k}} - E^M_{\mathbf{k}} = \frac{\beta}{b_0k^2}\operatorname{Re} H_{\mathbf{k}}. \tag{4.31}$$

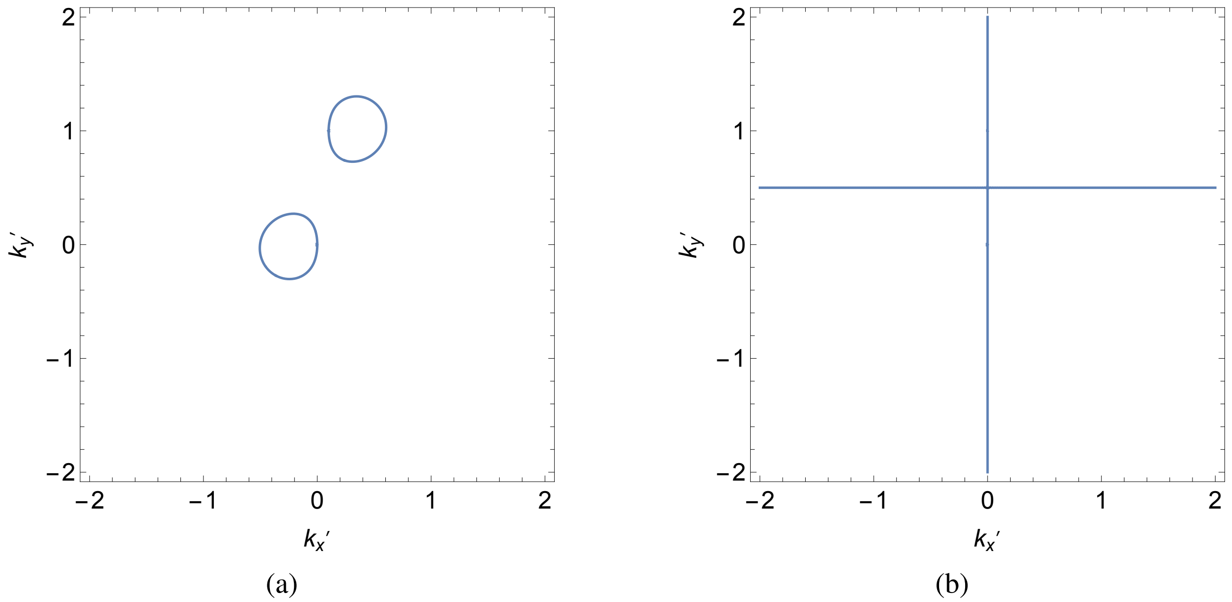


Figure 4.1: Plots of the resonance surfaces for the Rossby-Alfvén process $-,- \to +$, with $b_0 = 1$ and $\beta = 0.2$. In (a), we set $\mathbf{k} = (0.1, 1)$ and in (b), we set $\mathbf{k} = (0, 1)$.

Thus, at the level of weak turbulence, the stationary cross-helicity is *equivalent* to a deviation from energy equipartition, which also determines momentum transport. Another view is that the alignment of the velocity field with the magnetic field is determined by the alignment of the Elsässer populations. Moreover, one may rearrange this to identify the left-hand side with the Lorentz force and the right-hand side with the turbulent emf and obtain:

$$\frac{\langle \partial_t \nabla^2 \tilde{\psi} \rangle_{\mathbf{k}}}{\langle \partial_t \tilde{A} \rangle_{\mathbf{k}}} = k_{\mathrm{MR}}^2. \tag{4.32}$$

This tells us that the system organizes so that, at each scale k, the ratio of the rate of production of velocity fluctuations to the rate of production of magnetic fluctuations is k_{MR}^2/k^2, an explicit form of the maxim that at lengthscales larger than ℓ_{MR}, the turbulence is primarily kinetic, and at smaller scales it is primarily magnetic.

The vanishing of the cross-correlator also tells us that the imaginary part of the cross helicity vanishes, $\mathrm{Im}\, H_{\mathbf{k}} = 0$. This is equivalent to the statement that the turbulent flux of A vanishes in weak turbulence theory and consistent with the Zel'dovich theorem with sufficiently large b_0 and/or sufficiently small fluctuations $\langle \tilde{b}^2 \rangle$. We should not be surprised, as even a *weak* mean field is known to quench the turbulent resistivity [CV91, GD94, GD96, FDC19].

We note that these results still do not allow much progress with the spectral equations, since the amplitude of the oscillations of the cross-correlator, still undetermined, plays a major role in the nonlinear dynamics.

4.3 Simulation results

4.3.1 Stationary mean energies and cross-helicity

We now present simulation results for this system, on a periodic box of size $2\pi \times 2\pi$ with a 512×512 mesh. We used a pseudospectral code within the Dedalus framework [BVO$^+$20]. The vorticity was forced on an annulus in Fourier space, centered at $k_f = 32$ and with a width of 8. The forcing had a fixed energy injection rate of $\varepsilon = 10^{-2}$ and a correlation time of $\tau_c = 5 \times 10^{-3}$. All simulations had fixed $\nu = \eta = 10^{-4}$ and $b_0 = 2$, and β was varied over a broad range. Hyperviscous terms with coefficients 10^{-8} were included in each equation to improve the stability properties. The magnetic Reynolds number for these simulations ranged from $\mathrm{Rm} \sim 6000 - 15000$. After many dissipation times, $t \gg 1/(\eta k_f^2)$, the system reached a stationary state and the mean kinetic and magnetic energies, cross helicity, and zonal mean (kinetic) energy fraction $\mathrm{zmf} = \int dk_y \, |\tilde{v}(k_x = 0, k_y)|^2 / \int d^2\mathbf{k} \, |\tilde{v}(k_x, k_y)|^2$ were all recorded. A zmf exceeding ~ 0.5 signals an inverse cascade — most of the kinetic energy is at large parallel scales — and a transition to Rossby turbulence. These results are plotted in Fig. 4.2; in Fig. 4.3, the cross-helicity is compared to the predictions of § 4.2.1.

There is a wealth of information in these plots and it is worth discussing. When $\beta < b_0 k_f^2$, the magnetic and kinetic energies are nearly equal, suggesting that the turbulence is Alfvénized. As β grows larger, this equipartition begins to break down, and the magnetic energy drops, signaling the possible generation of a Reynolds stress. When β is significantly greater than $b_0 k_f^2$

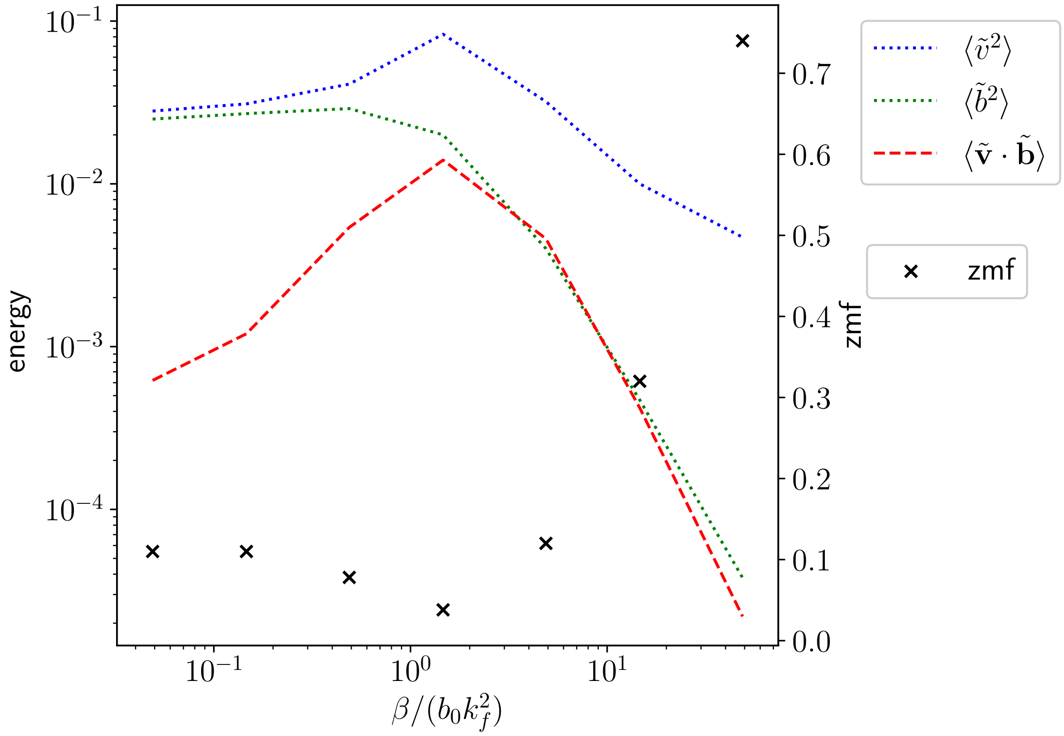


Figure 4.2: Plot of stationary turbulent energies, cross-helicity, and zmf obtained from simulation. All simulations were kinetically forced at an energy injection rate $\varepsilon = 10^{-2}$ at a scale $k_f = 32$ and had a mean field $\mathbf{b}_0 = 2\hat{x}$. The transition from Alfvénic to Rossby turbulence, signified by an imbalance between kinetic and magnetic energies and a large zmf, appears to start around a critical value $\beta = b_0 k_f^2$. This critical value, in general, will also depend on the resistivity [CD20]. The transition is presaged by an increase in the mean cross helicity, which peaks around the critical value. β is normalized in the plot by the critical value. The zmf reaches a minimum in the transitional regime, and becomes large (> 0.5) when $\beta \gg b_0 k_f^2$, signaling an inverse cascade.

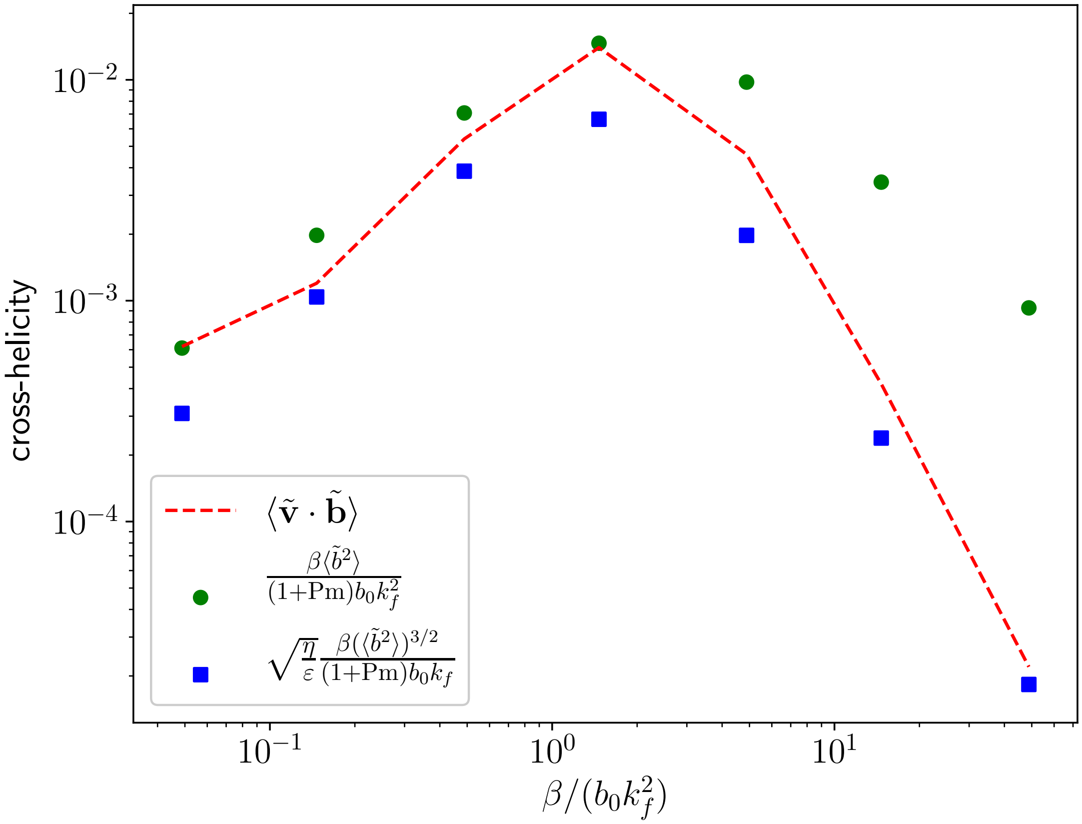


Figure 4.3: We again plot the cross-helicity from simulation (red dashed line) and compare it to the prediction from § 4.2.1. We use two estimates for the lengthscales: using $\ell_b = \ell_v = k_f^{-1}$ (green circles), which ceases to be good as β is large since the magnetic fluctuation scale becomes small, and a second estimate (blue squares) using a magnetic Taylor microscale $\ell_b = \sqrt{\eta/\varepsilon\langle\tilde{b}^2\rangle}$.

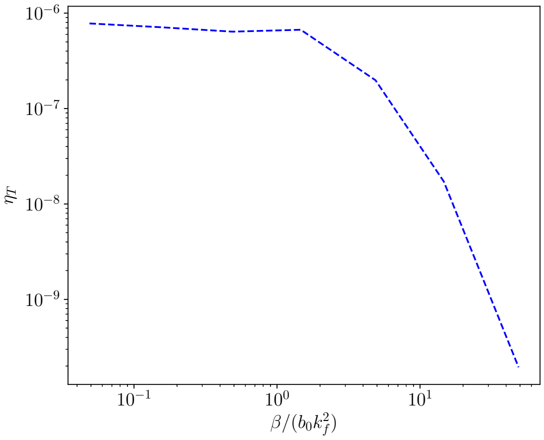


Figure 4.4: Plot of the turbulent resistivity $\eta_T = -\langle \tilde{v}_y \tilde{A} \rangle / b_0$ as a function of β.

most of the energy is almost entirely in the flow, and concentrated at large parallel lengthscales, signaling the inverse cascade and the transition to Rossby turbulence.

Meanwhile, as β nears the transition point, the cross-helicity grows, peaks near $\beta = b_0k_f^2$, and then drops again (that it tracks closely with the magnetic energy beyond this point appears to be coincidental to this choice of parameters). An estimate for the stationary cross-helicity using $\ell_b = \ell_v = k_f^{-1}$ begins to break down for large β; in this regime, the estimate $\ell_b = \sqrt{\frac{\eta}{\varepsilon}\langle \tilde{b}^2 \rangle}$ better fits the data.

Finally, we note the interesting fact that the zmf is nonmonotonic, and decreases during the transitional regime before becoming large.

4.3.2 Flux of magnetic potential

In Fig. 4.4 we plot the turbulent resistivity $\eta_T = -\langle \tilde{v}_y \tilde{A} \rangle / b_0$ as a function of β. It is generally very small, $\eta_T < \eta$, having been quenched by the strong mean field. This is consistent with the weak turbulence expectation $\operatorname{Im} H_{\mathbf{k}} = 0$. It drops precipitously when $\beta > b_0k_f^2$ due to the lack of magnetic activity.

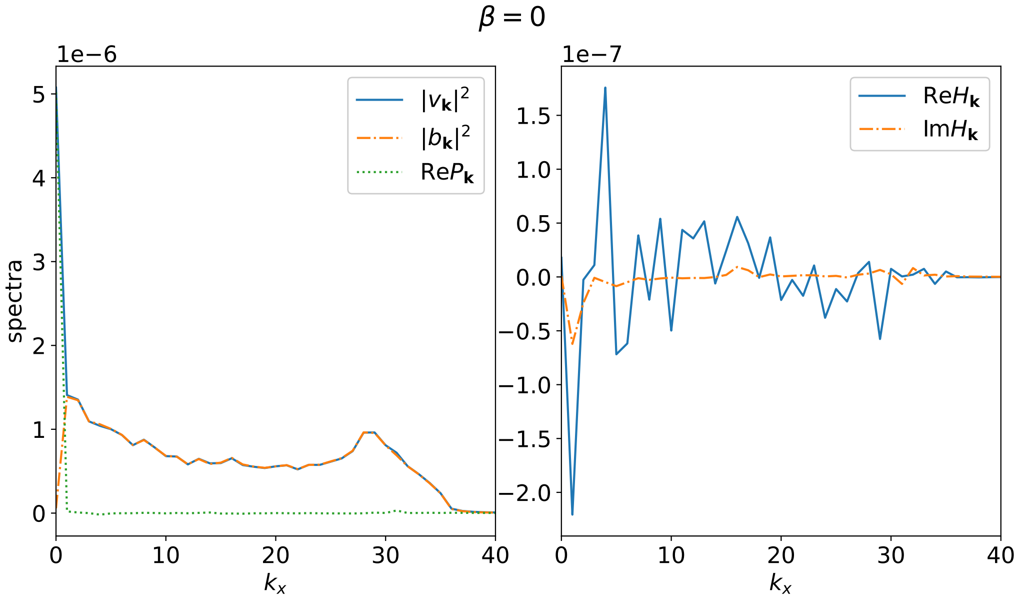


Figure 4.5: Stationary spectra, averaged over k_y, for $\beta = 0$. This is pure MHD, for the sake of comparison.

4.3.3 Spectra

We now turn our attention to the spectra. We present plots of the stationary energy spectra, cross-helicity spectra (including both real and imaginary parts), and Elsässer alignment $\mathrm{Re}\,\mathrm{P}_{\mathbf{k}} = \langle \tilde{v}_{\mathbf{k}}^2 \rangle - \langle \tilde{b}_{\mathbf{k}}^2 \rangle$, as functions of the parallel wavenumber k_x. The spectra are highly anisotropic; most of the interesting structure in the spectra are along k_x, whereas we found that the dependence on k_y largely reflects the structure of the forcing and is peaked at harmonics of k_f. Thus, we average the spectra over k_y, as well as late times (specifically, the final 25% of the timesteps of the simulations, which ran to $t = 400$).

The plots are shown in Figs. 4.6–4.8. The spectra are given at three values of β: 10^2, 3×10^3, and 10^5, respectively representing the small-β, transitional, and large-β regimes.

In the Alfvénic regime represented by $\beta = 10^2$, the kinetic and magnetic energy spectra are nearly identical except at small k_x. At the largest parallel scales ($k_x = 0$), the kinetic energy becomes large, and the magnetic energy goes to zero; thus the (real) alignment spectrum is basically a delta function centered at $k_x = 0$. At finite scales, there is an inertial range were the

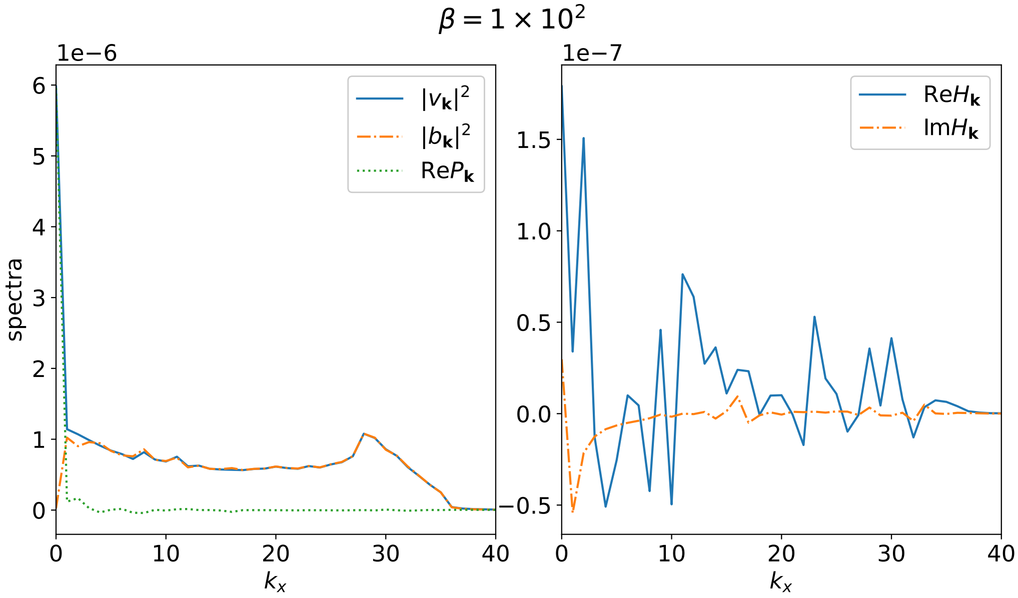


Figure 4.6: Stationary spectra, averaged over k_y, for $\beta = 10^2$.

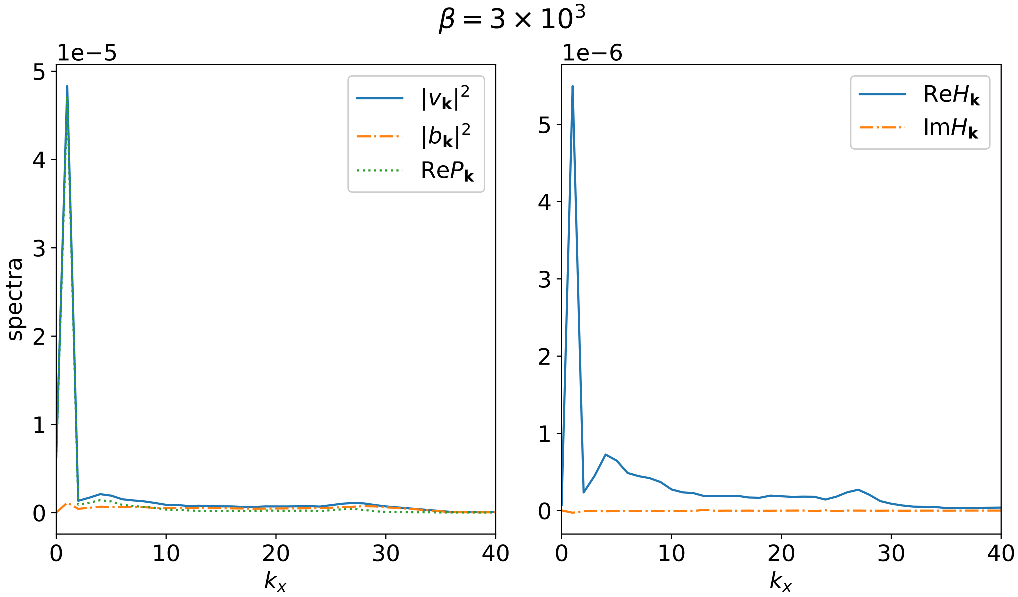


Figure 4.7: Stationary spectra, averaged over k_y, for $\beta = 3 \times 10^3$.

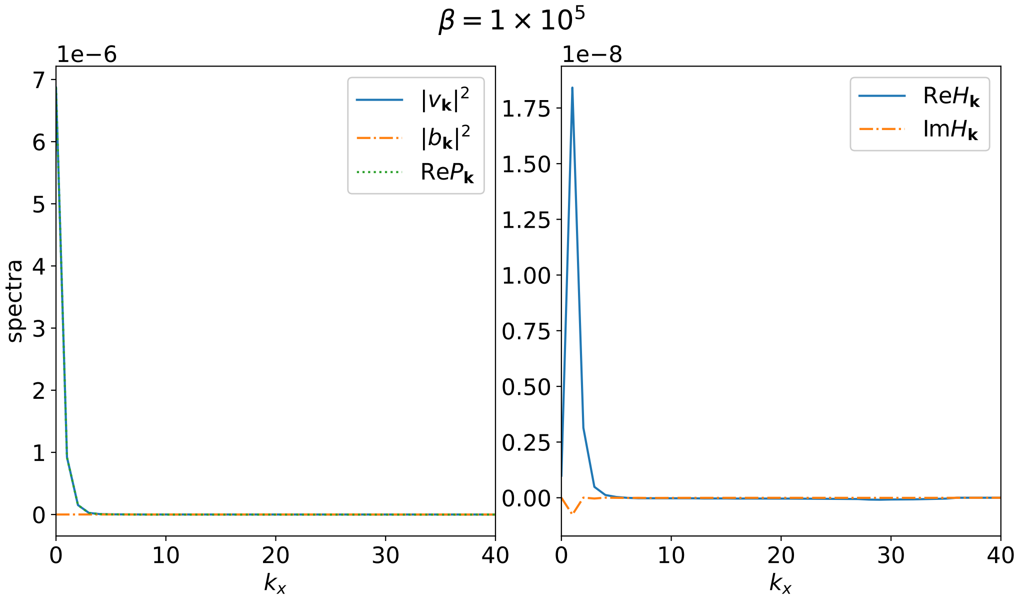


Figure 4.8: Stationary spectra, averaged over k_y, for $\beta = 10^5$.

energy spectra are essentially flat up to a cutoff, agreeing with weak turbulence theory for 2D MHD [TNG13]. A small real cross-helicity is mixed over a broad range of scales — this mixing occurs even when the total cross-helicity is zero.There is a small imaginary cross-helicity as well, concentrated at small but finite k_x, corresponding to the weak turbulent diffusivity. In Fig. 4.5, we also show corresponding results for $\beta = 0$, i.e. pure MHD, and conclude that the spectra are consistent.

In the transitional regime represented by $\beta = 3 \times 10^3$, the kinetic and magnetic energy spectra no longer agree at any scale, so the alignment is finite over a broad range. The peak in the kinetic energy and alignment spectra has shifted from $k_x = 0$ to a small but *finite* parallel lengthscale. The real cross-helicity, too, is peaked at this scale, and otherwise spread over a broad range. The imaginary cross-helicity is now negligible, consistent with the claim of the previous subsection.

Finally, in this Rossby-like regime represented by $\beta = 1 \times 10^5$, the magnetic energy is negligible, so the kinetic energy and alignment spectra are equivalent. We have checked that these decay like k_x^{-3}, consistent with the known theory of 2D kinetic turbulence [Kra67]. Both real and

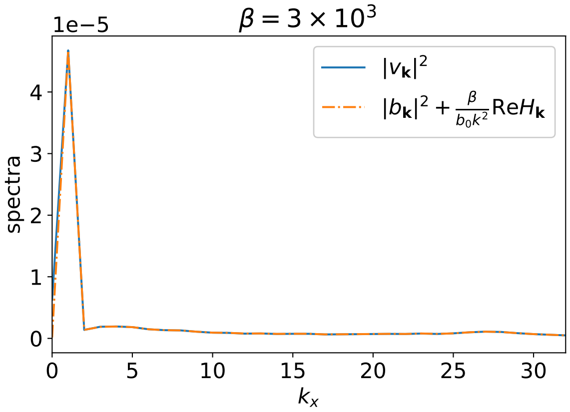


Figure 4.9: Comparison, for $\beta = 3\times10^3$, of kinetic energy spectra with weak turbulence estimate $|\tilde{b}_\mathbf{k}|^2 + \beta/(b_0k^2)\,\mathrm{Re}\,H_\mathbf{k}$.

imaginary parts of the cross helicity are small as there is very little magnetic activity.

The weak turbulence relationship Eq. 4.31 indeed holds to good approximation. In Fig. 4.9 we compare the kinetic energy spectra with $|\tilde{b}_\mathbf{k}|^2 + \beta/(b_0k^2)\,\mathrm{Re}\,H_\mathbf{k}$ for the case $\beta = 3\times10^3$; the curves lie on top of one another except at $k_x = 0$.

4.4 Discussion

We have studied the stationary spectra of a simple model of the solar tachocline, β-plane MHD turbulence, with a strong mean field. The results show three distinct regimes: an Alfvénic regime at small β, a Rossby regime at large β, and a transitional regime near $\beta \simeq b_0k_f^2$.

We have shown that the turbulent cross-helicity builds up in this system to a predictable level, and, moreover, the stationary cross-helicity spectra are entirely equivalent to the Maxwell-Reynolds stress differential in weak turbulence theory. Thus, momentum transport, as well as the energy partition, is set by the cross-helicity. This result is a consequence of the fact that the cross-correlator between the eigenmodes of the system must oscillate on a (fast) linear timescale.

In the strong turbulence regime, on the other hand, this relationship will break down; indeed, the very notion of linear Rossby-Alfvén modes ceases to be meaningful. However, we still anticipate that cross-helicity may play an important role, as the stationary cross-helicity is proportional to $\langle \tilde{b}^2 \rangle / b_0$, which may be very large in a strongly turbulent regime. In particular, there may be an effect on the flux of magnetic potential, which is essentially zero in weak turbulence, but has magnetic Reynolds number dependence [CV91, GD94, GD96] and intermittent spatial structure [FDC19] when the mean field is weak. Future research should investigate the strong turbulence regime, and determine how the turbulent magnetic diffusivity depends on β.

We have noted the general connection between cross-helicity and dynamo action, but not yet addressed the turbulent emf for the present system. While the interaction of the planetary vorticity gradient and the mean field suffice to generate a cross-helicity, in the present setup, $y \to -y$ reflection symmetry enforces $\int d^2\mathbf{x} \frac{k_x k_y}{k^2} \operatorname{Re} H_{\mathbf{k}} = 0$, so there is no mean emf. However, this might change if this symmetry is broken, say by a component of the mean field along $\hat{y}$ or an inhomogeneity (such as a shear flow or a magnetic field gradient). It would be interesting to study transport in such a system.

Of course, there is no dynamo in two dimensions, so any such study would constitute no more than a toy model for emf generation through cross-helicity. Thus, we suggest studying an extension of the system to three dimensions. Here, cross-helicity has a topological interpretation (see, for example, [Mof78]) as a measure of the mutual linkage between flux tubes and vortex tubes, so the effect of cross helicity may be more impactful than in 2D, where such an interpretation ceases to make sense. We also have in 3D [Mof78]

$$\langle (\tilde{\mathbf{v}} \cdot \tilde{\mathbf{b}})^2 \rangle + \langle (\tilde{\mathbf{v}} \times \tilde{\mathbf{b}})^2 \rangle = \langle (\tilde{v}\tilde{b})^2 \rangle, \tag{4.33}$$

which suggests that the cross-helicity may inhibit dynamo action.

Chapter 4, in full, is currently being prepared for submission for publication of the material. R. A. Heinonen and P. H. Diamond (2021). The dissertation author was the primary investigator and author of this material.

Chapter 5

Conclusion

We have investigated turbulent transport in plasma systems, relevant to fusion in both manmade devices and in the natural world. In all the problems we have studied, the nonlinear interaction between microscopic turbulence lengthscales (in the tokamak, usually ρ_s, and in MHD, the forcing scale) conspires to generate dynamics on a much larger lengthscale, such a gradient scale length, the system size, or some combination thereof. The results can be extremely important to the behavior of the system: avalanches can deliver large fractions of the flux; zonal flows regulate turbulence and enable improved confinement. The interaction of avalanches (and other forms of turbulence spreading) with zonal flow structures, which we will discuss later in this conclusion, is also of fundamental interest, though it was not investigated in this dissertation.

The first project contained in this dissertation used deep learning to identify the nonlinear dynamics of the drift-wave/zonal flow system. This system includes a complicated feedback loop between a dynamical gradient, which excites turbulence beyond a threshold; the turbulence, which tends to relax the gradient and spawns zonal flow; and the zonal flows, which suppress the turbulence. Deep learning recovered these relationships (by modeling the fluxes that govern them) and also drew attention to a far less well-studied one, the direct interaction between profile and flow through the vorticity gradient. This interaction, ultimately a finite gyroradius effect, is a simple route to the formation of a staircase pattern in the profile — an alternative, or perhaps partner, to mechanisms based on bistability. The deep learning approach also showed that zonal flow formation may be modeled as a Cahn-Hilliard equation in one dimension, a conservative model where the Reynolds stress is given by a negative viscosity, regularized by hyperviscous terms and more complicated factors involving the profile gradients. These latter factors are especially difficult to calculate by hand.

One must acknowledge, on the other hand, that much of the basic structure of the learned models could be supposed beforehand through a combination of symmetries and quasilinear modeling. Indeed, the application of physical intuition was a necessary ingredient for successful training. We were also confined to a regime with robust zonal flows, which required the gradient drive to not be too strong. Otherwise, the dynamics were largely controlled by the motion and interactions of strong, localized vortices, so that a one-dimensional mean field model was completely unreasonable.

We feel that the utility of the deep learning approach to modeling would be greatest in more complex systems than the adiabatic Hasegawa-Wakatani system studied here, where perhaps some physical intuition exists, but the requisite calculations are very demanding. These might include multi-channel plasma systems, say with temperature and magnetic dynamics, but where at least one degree of freedom in space still has special significance — the radial direction, in the tokamak. We can also imagine increasing the complexity of the machine learning model (by, say, relaxing the locality assumption), though we stress this will necessarily harm interpretability.

An early goal of the deep learning research was to study the interplay of turbulence spreading (the subject of Ch. 3) and zonal flow dynamics. Other studies (for example, [AD17, AD16, SD20]) have suggested that a certain degree of spreading is a necessary ingredient for inhomogeneous PV mixing, but too much spreading can destroy zonal flow structures. Zonal flows may also regulate spreading and limit its spatial extent through eddy shearing. However, interestingly, spreading resisted our attempts to model it with the deep learning approach and integrate it with the mean-field dynamics of the zonal flow and the profile. In spite of the fact that turbulence spreading is frequently modeled using mean-field theory — with the intensity flux scaling with both the mesoscale intensity gradient and the intensity itself — our efforts to train such a model for the flux of turbulent potential enstrophy weren?t successful. It could be that spreading plays a limited role in the regime of HW turbulence that we studied (certainly intensity dynamics are important in the more strongly-driven regime where vortex dynamics predominate), or that spreading in HW fails to adhere to a model that is local in time and/or space. More research is required to be sure of this.

The spreading model of Ch. 3 nevertheless made this assumption of spreading by turbulent diffusion. Regardless of the precise "correctness" of the expression for the intensity flux, when combined with a bistable model for the turbulence intensity, it manages to reproduce some of the basic features of spreading dynamics, in addition to a plausible mechanism for avalanching. We emphasize that that this mechanism, triggered by the exceedance of a nonlinear instability threshold, may be testable in laboratory experiment. A possible setup would be one similar to [VCMMS15] where the plasma is locally excited by a source, and the spatiotemporal response is observed. Care must be taken to distinguish the response from that of a linear mode. Our prediction that subcritical turbulence spreading should penetrate a (weakly) stable region *ballistically* is also testable, and indeed, this behavior was possibly observed in gyrokinetic simulations with magnetic shear [YKDH14]. This should be confirmed with additional numerical studies. Finally, we also note the PDE under study is of fundamental interest for a variety of applications, and most of our conclusions, including the avalanche threshold, are universal properties of the system.

Lastly, we studied momentum transport in beta-plane MHD turbulence, with a focus on cross-helicity. Our key results were an explicit expression for the stationary cross-helicity, and a direct relationship, valid in weak turbulence theory, between the time-averaged cross-helicity spectrum and Elsässer alignment spectrum (which determines momentum transport). This relationship also makes explicit the assertion of [DIIS07] that the nature of the turbulent fluctuations — kinetic or magnetic — is separated according to lengthscale, with the critical separation scale being the magnetic Rhines scale.

We primarily took the approach of weak turbulence for the sake of analytical tractability, but even the weak turbulence spectral equations resisted our attempts at explicit solution. Our hope was to treat the planetary vorticity gradient β as a perturbation to weak 2D MHD turbulence and to determine the modulational response to this perturbation. However, we found that the perturbation is singular. In a real system, the $\beta \to 0$ singularity is regularized by the presence of a nonlinear resonance broadening, so that the transition from β-plane turbulence to 2D MHD is smooth, but this is not helpful for analytical calculation. More research should be done on this problem in the strong turbulence regime, where the cross-helicity could be very large, as well as in three dimensions (or perhaps 2.5 dimensions, akin to a shallow-water system), where cross-helicity has special significance as a topological quantity, and dynamo action is possible.

Finally, we note that β-plane MHD shares several features in common with HW, including (in the high-β regime) zonal flow formation by Reynolds stress. If the vorticity equation is slightly modified ($\partial_t \nabla^2 \phi \to \partial_t (\nabla^2 \phi - \phi)$), it also serves as a simple model for the tokamak with perpendicular magnetic fluctuations (drift-Alfvén turbulence), which suggests that an excess of MHD activity destroys zonal flow structures and limits access to the H-mode.

One may naturally wonder if the deep learning approach of Ch. 2 could be applied to the β-plane MHD system. A key difference is that MHD lacks the adiabatic relationship between the two fields (n and ϕ in HW, A and ψ in MHD) which we exploited in HW to reduce the number of parameters which determine the flux. In MHD, one would likely need 3 independent turbulence intensities, representing magnetic fluctuations, kinetic fluctuations, and correlations between the two. Moreover, outside of the high-β regime, there is no zonal symmetry, so one would not be able to reduce the number of spatial dimensions for arbitrary β; instead, one could imagine the "mean fields" being local averages in 2D windows rather than local averages. Both of these issues are also present in *nonadiabatic* HW and would need to be contended with if one wished to extend the model to arbitrary α.

Statistician George Box wrote, “All models are wrong, but some models are useful.” Certainly, plasma turbulence is an enormously complicated problem, and the simple mathematical models we studied in this dissertation are necessarily incomplete descriptions. However, therein also lies their utility, for the simplicity permits tractable analysis and better physics understanding. And, conversely, if one cannot understand the simple models, we have no hope of understanding complicated ones.

Appendix A

The Taylor identity

The Taylor identity (named for G. I. Taylor) states that, in a periodic system, the vorticity flux is equivalent to the gradient of the Reynolds stress (that is, the Reynolds force). We show this briefly. The vorticity flux is given (up to a sign) by

$$\langle \partial_y \tilde{\phi} \nabla_\perp^2 \tilde{\phi} \rangle = \langle \partial_y \tilde{\phi} \partial_x^2 \tilde{\phi} \rangle + \langle \partial_y \tilde{\phi} \partial_y^2 \tilde{\phi} \rangle. \tag{A.1}$$

The first term may be rewritten

$$\langle \partial_y \tilde{\phi} \partial_x^2 \tilde{\phi} \rangle = \left\langle \partial_x (\partial_x \tilde{\phi} \partial_y \tilde{\phi}) - \frac{1}{2} \partial_y \left((\partial_x \tilde{\phi})^2 \right) \right\rangle = \partial_x \langle \partial_x \tilde{\phi} \partial_y \tilde{\phi} \rangle, \tag{A.2}$$

where we have used periodicity in the y-direction. The other term may be rewritten

$$\langle \partial_y \tilde{\phi} \partial_y^2 \tilde{\phi} \rangle = \left\langle \frac{1}{2} \partial_y \left((\partial_y \tilde{\phi})^2 \right) \right\rangle = 0, \tag{A.3}$$

again using periodicity. We are left with

$$\langle \partial_y \tilde{\phi} \nabla_\perp^2 \tilde{\phi} \rangle = \partial_x \langle \partial_x \tilde{\phi} \partial_y \tilde{\phi} \rangle. \tag{A.4}$$

The RHS is the Reynolds force, as claimed.

Appendix B

Error quantification for deep learning

We attempt to quantify the accuracy of the DNN models in two ways.

First, a typical final validation loss (excluding the regularization term) for a trained DNN model was ~ 0.004 for the Reynolds stress and ~ 0.003 for the particle flux. If we invert the these values for the logcosh, we obtain rough estimates for a typical error term: $\Delta\Pi \sim 0.09$ and $\Delta\Gamma \sim 0.07$. Typical values for the predicted fluxes are $|\Pi| \leq 0.3$ and $|\Gamma| \leq 3$, so this estimate indicates the error term is more significant for the Reynolds stress.

A second estimate for the error comes from the variance among the ensemble of ten DNNs. This error is illustrated in Figs. B.1–B.3. It is less clear from this picture that the uncertainty in the Reynolds stress is more significant. A couple interesting features are apparent: first, there is a peak in the standard deviation of Γ near $N' = U' = 0$ when $\varepsilon > 0$. This is likely associated with this condition not being easily realized in our simulation; the DNN is generalizing from the simulation data to predict this point and its neighborhood. Similarly, there is a peak near $U = 0$ in the standard deviation of Π. We note that $U = 0$ tends to correlate with small ε, so again, simultaneous $U = 0$ and $\varepsilon > 0$ is not easily realized. The uncertainty also generally scales with the turbulence intensity, reflecting the fact that both the flux itself and the noise signal scale with intensity.

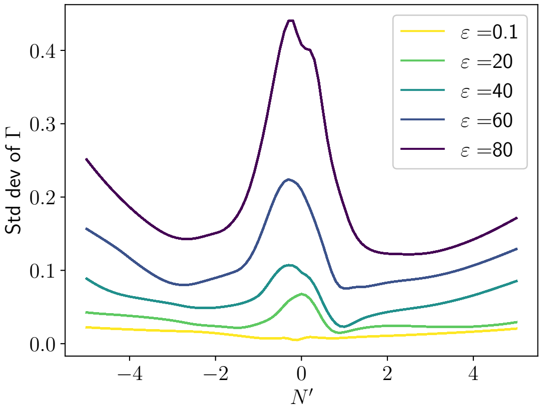


Figure B.1: Plot of the standard deviation among the ensemble of DNN models for the diffusive/diagonal part of the particle flux ($U = U' = U'' = 0$).

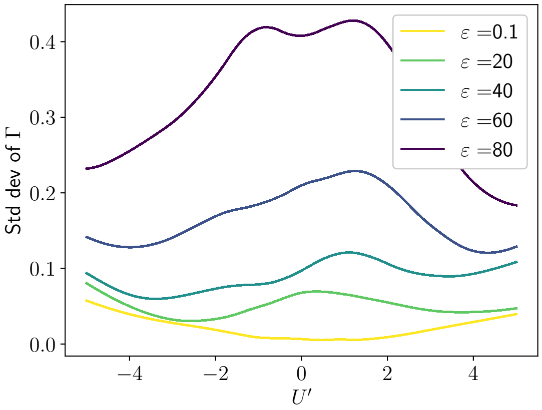


Figure B.2: Plot of the standard deviation among the ensemble of DNN models for the nondiffusive/off-diagonal part of the particle flux ($U = U'' = N' = 0$).

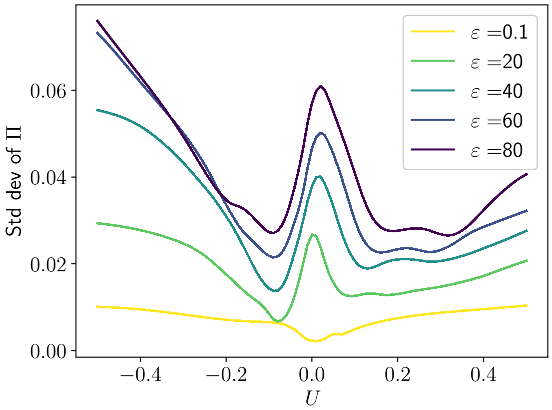


Figure B.3: Plot of the standard deviation among the ensemble of DNN models for the Reynolds stress, when $N' = 2$ and $U' = U'' = 0$.

Appendix C

Reduction of Gil-Sornette model to bistable model

Let us rewrite the Gil-Sornette model in an equivalent form which makes the analogy to our model explicit:

$$\partial_t I = \gamma_1 \left(\frac{|\partial_x T|}{g_c} - 1 \right) I + \gamma_2 I^2 - \gamma_3 I^3 + \partial_x (D_1 I \partial_x I) \tag{C.1}$$

$$\partial_t T = \partial_x (D_2 I \partial_x T). \tag{C.2}$$

Note the addition of a spreading term. Clearly, in the limit that the dynamics governing I are faster than those governing T, $\partial_x T$ behaves as a constant in Eq. (C.1), and our model may be applied.

In the opposite limit where I is slaved to T, the steady state local gradient is $\partial_x T \propto 1/I$, where the constant of proportionality depends on boundary conditions. Plugging this into Eq. (C.1), the linear term becomes $c - \gamma_1 I$ for some constant c. If c is zero, we exactly recover the linearly damped case of our model. Otherwise, it represents a generalization which includes a turbulence source/sink. Analysis of such a model is more complicated, but the qualitative features should be quite similar.

Appendix D

Variational form of Gil-Sornette model

One naturally wonders if the sandpile model of Gil and Sornette can be expressed, in analogy to Eq. (3.21), in terms of the (monotonic) dissipation of a Lyapunov functional. This does not appear to be precisely the case, but instead we may write

$$\partial_t S = \gamma \frac{\delta \mathscr{F}}{\delta S} \tag{D.1}$$

$$\partial_t h = -D(\partial_x h_c)^2 \frac{\delta \mathscr{F}}{\delta h} \tag{D.2}$$

with

$$\mathscr{F} = \int dx \left[\frac{1}{2} \left(\frac{(\partial_x h)^2}{(\partial_x h_c)^2} - 1 \right) S^2 + \frac{1}{4} \alpha S^4 - \frac{1}{6} S^6 \right], \tag{D.3}$$

The asymmetry in the signs in Eqs. (D.1) and (D.2) is crucial. Rather than monotonic evolution of $\mathscr{F}$, we have

$$\frac{d\mathscr{F}}{dt} = \int dx \left[\frac{1}{\gamma} (\partial_t S)^2 - \frac{1}{D(\partial_x h_c)^2} (\partial_t h)^2 \right]. \tag{D.4}$$

A competition then becomes apparent: at fixed S, the h dynamics works to minimize the gradient, and at a fixed h profile, the S dynamics works to drive the system to a fixed point of the quintic polynomial. These trends are in mutual opposition. One can thus expect at least two regimes with quite different dynamics, based on the relative magnitudes of the timescales $(D(\partial_x h_c)^2)^{-1}$ and

γ^{-1}.

Appendix E

Standing waves in bistable model

A brief discussion of the standing wave states of the model follows. We seek solutions $I(x)$ to

$$\begin{aligned} 0 &= \partial_x(D(I)\partial_x I) + I(1-I)(I-\alpha) \\ &= D(I)I'' + D'(I)(I')^2 + I(1-I)(I-\alpha) \end{aligned} \tag{E.1}$$

such that $I(\pm\infty, t) = 0$. Writing $F(I) = (I')^2$ (so that $I'' = \frac{1}{2}F'(I)$) and reducing the order yields

$$D(I)^2 F(I) = I^2 I'^2 = 2I^3 \left(\frac{\alpha}{3} - \frac{1+\alpha}{4} I + \frac{I^2}{5} \right) + C. \tag{E.2}$$

The parameter C sets the behavior of the standing wave when I vanishes; we require $C \geq 0$.

We can find the closed-form solution if $C = 0$, i.e. for a wave with nonsingular gradient at $I = 0$. Integrating directly, one finds [BF71]

$$I_0(x) = \beta_- \operatorname{sn}^2 \left(\sqrt{\frac{\beta_+}{10}} (x - x_0), \sqrt{\frac{\beta_-}{\beta_+}} \right) \tag{E.3}$$

for arbitrary x_0, where

$$\beta_\pm = \frac{5}{8}\left(\alpha + 1 \pm \sqrt{1 - \frac{34}{15}\alpha + \alpha^2}\right). \tag{E.4}$$

Here, sn is the Jacobi elliptic function. This wave is periodic with period $2\sqrt{\frac{10}{\beta_+}}K\left(\sqrt{\frac{\beta_-}{\beta_+}}\right)$, where K is the complete elliptic integral of the first kind, but we can restrict the support to just one period so that it has a single "hump." It is even with appropriate choice of x_0.

The standing waves are linearly unstable to perturbations, as can be shown by the following argument. Write $I(x,t) = \bar{I}(x) + \tilde{I}(x,t)$, where $\bar{I}$ is a solution to Eq. (E.1), $\tilde{I}$ is assumed to be small, substitute into Eq. (E.1), and linearize to obtain

$$\partial_t \tilde{I} = \partial_{xx}\left[D(\bar{I})\tilde{I}\right] + \tilde{I}f'(\bar{I}). \tag{E.5}$$

Now, make the ansatz $\tilde{I}(x,t) = \psi(x)e^{-\lambda t}$, which transforms Eq. (E.5) into

$$-\lambda\psi = \partial_{xx}\left[D(\bar{I})\psi\right] + \psi f'(\bar{I}), \tag{E.6}$$

which is a Sturm-Liouville equation for $D(\bar{I})\psi$. By translational symmetry (or direct substitution), $\psi_0 = \partial_x \bar{I}$ is a solution with zero eigenvalue. Provided $\bar{I}$ is non-monotonic as it is for I_0 above, ψ_0 has at least one zero, so according to the theorems of Sturm-Liouville theory, there must be another eigenfunction with no zeros and with eigenvalue $\lambda < 0$, implying instability.

Appendix F

Derivation of weak turbulence spectral equations

Begin by defining the mode amplitudes $\hat{\phi}_{\mathbf{k}}^{\alpha} = e^{i\omega_{\mathbf{k}}^{\alpha} t}\phi_{\mathbf{k}}^{\alpha}$. Equation (4.18) becomes

$$\partial_t \hat{\phi}_{\mathbf{k}}^{\alpha} = \frac{1}{2}\sum_{\beta\gamma}\sum_{\mathbf{k}'+\mathbf{k}''=\mathbf{k}} e^{i(\omega_{\mathbf{k}}^{\alpha}-\omega_{\mathbf{k}'}^{\beta}-\omega_{\mathbf{k}''}^{\gamma})t} M_{\mathbf{k},\mathbf{k}',\mathbf{k}''}^{\alpha\beta\gamma}\hat{\phi}_{\mathbf{k}'}^{\beta}\hat{\phi}_{\mathbf{k}''}^{\gamma}. \tag{F.1}$$

We proceed via time-dependent perturbation theory, letting

$$\hat{\phi}_{\mathbf{k}}^{\alpha}(t) = \hat{\phi}_{\mathbf{k}}^{\alpha}(0) + \delta\hat{\phi}_{\mathbf{k}}^{\alpha,(1)}(t) + \delta\hat{\phi}_{\mathbf{k}}^{\alpha,(2)}(t) + \dots. \tag{F.2}$$

Then

$$\delta\hat{\phi}_{\mathbf{k}}^{\alpha,(1)}(t) = \frac{1}{2}\sum_{\mathbf{k}'+\mathbf{k}''=\mathbf{k}}\sum_{\beta\gamma} M_{\mathbf{k},\mathbf{k}',\mathbf{k}''}^{\alpha\beta\gamma}\hat{\phi}_{\mathbf{k}'}^{\beta}(0)\hat{\phi}_{\mathbf{k}''}^{\gamma}(0)\int_0^t dt'\, e^{i(\omega_{\mathbf{k}}^{\alpha}-\omega_{\mathbf{k}'}^{\beta}-\omega_{\mathbf{k}''}^{\gamma})t'} \tag{F.3}$$

and

$$\begin{aligned}\delta\hat{\phi}_{\mathbf{k}}^{\alpha,(2)}(t) = \frac{1}{2}\sum_{\beta\gamma\beta'\gamma'}\sum_{\substack{\mathbf{k}'+\mathbf{k}''=\mathbf{k}\\ \mathbf{q}'+\mathbf{q}''=\mathbf{k}'}} & M^{\alpha\beta\gamma}_{\mathbf{k},\mathbf{k}',\mathbf{k}''}M^{\beta\beta'\gamma'}_{\mathbf{k}',\mathbf{q}',\mathbf{q}''}\hat{\phi}^{\gamma}_{\mathbf{k}''}(0)\hat{\phi}^{\beta'}_{\mathbf{q}'}(0)\hat{\phi}^{\gamma'}_{\mathbf{q}''}(0)\\ & \times\int_0^t dt' e^{i(\omega^{\alpha}_{\mathbf{k}}-\omega^{\beta}_{\mathbf{k}'}-\omega^{\gamma}_{\mathbf{k}''})t'}\int_0^{t'} dt''\, e^{i(\omega^{\beta}_{\mathbf{k}'}-\omega^{\beta'}_{\mathbf{q}'}-\omega^{\gamma'}_{\mathbf{q}''})t''},\end{aligned}\tag{F.4}$$

where we have combined terms by exchanging species indices and using the symmetry of the coupling coefficients.

We are interested in the evolution of $C^{\alpha\alpha'}_{\mathbf{k}}(t) \equiv \langle\hat{\phi}^{\alpha}_{\mathbf{k}}\hat{\phi}^{\alpha'*}_{\mathbf{k}}\rangle$. Working to second order,

$$\Delta C^{\alpha\alpha'}_{\mathbf{k}} = C^{\alpha\alpha'}_{\mathbf{k}}(t) - C^{\alpha\alpha'}_{\mathbf{k}}(0) = \langle\delta\hat{\phi}^{\alpha,(1)}_{\mathbf{k}}\delta\hat{\phi}^{\alpha',(1)*}_{\mathbf{k}}\rangle + \langle\delta\hat{\phi}^{\alpha,(2)}_{\mathbf{k}}\hat{\phi}^{\alpha'*}_{\mathbf{k}}(0) + \hat{\phi}^{\alpha}_{\mathbf{k}}(0)\delta\hat{\phi}^{\alpha',(2)*}_{\mathbf{k}}\rangle + \ldots,\tag{F.5}$$

where we have anticipated that the first-order terms make no contribution.

Now,

$$\begin{aligned}\Delta C^{\alpha\alpha',(1)}_{\mathbf{k}} &= \langle\delta\hat{\phi}^{\alpha,(1)}_{\mathbf{k}}\delta\hat{\phi}^{\alpha',(1)*}_{\mathbf{k}}\rangle\\ &= \frac{1}{4}\sum_{\beta\gamma\beta'\gamma'}\sum_{\substack{\mathbf{k}'+\mathbf{k}''=\mathbf{k}\\ \mathbf{q}'+\mathbf{q}''=\mathbf{k}}} M^{\alpha\beta\gamma}_{\mathbf{k},\mathbf{k}',\mathbf{k}''}M^{\alpha'\beta'\gamma'*}_{\mathbf{k},\mathbf{q}',\mathbf{q}''}\langle\hat{\phi}^{\beta}_{\mathbf{k}'}(0)\hat{\phi}^{\gamma}_{\mathbf{k}''}(0)\hat{\phi}^{\beta'*}_{\mathbf{q}'}(0)\hat{\phi}^{\gamma'*}_{\mathbf{q}''}(0)\rangle\\ &\quad\times\int_0^t dt'\int_0^t dt''\, e^{i(\omega^{\alpha}_{\mathbf{k}}-\omega^{\beta}_{\mathbf{k}'}-\omega^{\gamma}_{\mathbf{k}''})t'}e^{-i(\omega^{\alpha'}_{\mathbf{k}}-\omega^{\beta'}_{\mathbf{q}'}-\omega^{\gamma'}_{\mathbf{q}''})t''}.\end{aligned}\tag{F.6}$$

Next, we make the random phase approximation, which means we can apply Wick's theorem to the four-mode functions, and assume spatial homogeneity, giving

$$\begin{aligned}\langle\hat{\phi}^{\beta}_{\mathbf{k}'}(0)\hat{\phi}^{\gamma}_{\mathbf{k}''}(0)\hat{\phi}^{\beta'*}_{\mathbf{q}'}(0)\hat{\phi}^{\gamma'*}_{\mathbf{q}''}(0)\rangle &= \langle\hat{\phi}^{\beta}_{\mathbf{k}'}(0)\hat{\phi}^{\beta'*}_{\mathbf{k}'}(0)\rangle\langle\hat{\phi}^{\gamma}_{\mathbf{k}''}(0)\hat{\phi}^{\gamma'*}_{\mathbf{k}''}(0)\rangle\delta_{\mathbf{k}'\mathbf{q}'}\delta_{\mathbf{k}''\mathbf{q}''}\\ &\quad+\langle\hat{\phi}^{\beta}_{\mathbf{k}'}(0)\hat{\phi}^{\gamma'*}_{\mathbf{k}'}(0)\rangle\langle\hat{\phi}^{\gamma}_{\mathbf{k}''}(0)\hat{\phi}^{\beta'*}_{\mathbf{k}''}(0)\rangle\delta_{\mathbf{k}'\mathbf{q}''}\delta_{\mathbf{k}''\mathbf{q}'}\end{aligned}\tag{F.7}$$

and allowing us to make several simplifications.

We need to evaluate the integral

$$I_1 = \int_0^t dt' \int_0^t dt''\, e^{i\Delta\omega t'} e^{-i\Delta\omega' t''} = \frac{4e^{i(\Delta\omega-\Delta\omega')t/2}\sin(\Delta\omega t/2)\sin(\Delta\omega' t/2)}{\Delta\omega\Delta\omega'} \tag{F.8}$$

where $\Delta\omega = \omega_{\mathbf{k}}^{\alpha} - \omega_{\mathbf{k}'}^{\beta} - \omega_{\mathbf{k}''}^{\gamma}$ and $\Delta\omega' = \omega_{\mathbf{k}}^{\alpha'} - \omega_{\mathbf{k}'}^{\beta'} - \omega_{\mathbf{k}''}^{\gamma'}$.

We seek the limit of this integral, in a distributional sense, as $t \to \infty$. We only keep terms linear in t; physically, we are interested in times $\omega^{-1} < t < \gamma_{NL}^{-1}$. The long-time limit vanishes unless $\Delta\omega = \Delta\omega' = 0$, but its value depends on whether we take the $\Delta\omega' \to \Delta\omega$ limit or $t \to \infty$ limit first. The sensible choice is the former, and we obtain

$$I_1(t\to\infty) = 2\pi t\,\delta(\Delta\omega)\,\delta_{\Delta\omega,\Delta\omega'}. \tag{F.9}$$

The support of the Kronecker delta has measure zero in $(\mathbf{k}', \mathbf{k}'')$ space unless $\Delta\omega = \Delta\omega'$ *identically*, i.e. $\alpha = \alpha'$, $\beta = \beta'$, $\gamma = \gamma'$. Using these simplifications, we obtain

$$\Delta C_{\mathbf{k}}^{\alpha\alpha',(1)} = \pi t \sum_{\mathbf{k}'+\mathbf{k}''=\mathbf{k}} \sum_{\beta\gamma} |M_{\mathbf{k},\mathbf{k}',\mathbf{k}''}^{\alpha\beta\gamma}|^2 C_{\mathbf{k}'}^{\beta\beta} C_{\mathbf{k}''}^{\gamma\gamma} \delta(\omega_{\mathbf{k}}^{\alpha} - \omega_{\mathbf{k}'}^{\beta} - \omega_{\mathbf{k}''}^{\gamma})\,\delta_{\alpha\alpha'}, \tag{F.10}$$

and, applying similar reasoning to the other terms of $\Delta C_{\mathbf{k}}^{\alpha\alpha'}$,

$$\begin{aligned}\Delta C_{\mathbf{k}}^{\alpha\alpha',(2)} &= \langle \delta\hat{\phi}_{\mathbf{k}}^{\alpha,(2)} \hat{\phi}_{\mathbf{k}}^{\alpha'*}(0) + \hat{\phi}_{\mathbf{k}}^{\alpha}(0)\,\delta\hat{\phi}_{\mathbf{k}}^{\alpha',(2)*} \rangle \\ &= t \sum_{\mathbf{k}'+\mathbf{k}''=\mathbf{k}} \sum_{\beta\gamma} M_{\mathbf{k},\mathbf{k}',\mathbf{k}''}^{\alpha\beta\gamma} M_{\mathbf{k}',\mathbf{k},-\mathbf{k}''}^{\beta\alpha\gamma} C_{\mathbf{k}}^{\alpha\alpha'} C_{\mathbf{k}''}^{\gamma\gamma} \left(\pi\delta(\omega_{\mathbf{k}}^{\alpha} - \omega_{\mathbf{k}'}^{\beta} - \omega_{\mathbf{k}''}^{\gamma}) + i\mathscr{P}\frac{1}{\omega_{\mathbf{k}}^{\alpha} - \omega_{\mathbf{k}'}^{\beta} - \omega_{\mathbf{k}''}^{\gamma}} \right) \\ &\quad + \text{c.c.}', \end{aligned} \tag{F.11}$$

where c.c.$'$ means the complex conjugate with $\alpha \leftrightarrow \alpha'$.

Finally, we obtain the claimed result by combining $\Delta C_{\mathbf{k}}^{\alpha\alpha',(1)}$ with $\Delta C_{\mathbf{k}}^{\alpha\alpha',(2)}$ and approximating

$$\partial_t C_{\mathbf{k}}^{\alpha\alpha'} \simeq \frac{\Delta C_{\mathbf{k}}^{\alpha\alpha'}}{t}.$$

Bibliography

[AD16] Arash Ashourvan and PH Diamond. How mesoscopic staircases condense to macroscopic barriers in confined plasma turbulence. *Physical Review E*, 94(5):051202, 2016.

[AD17] Arash Ashourvan and Patrick H Diamond. On the emergence of macroscopic transport barriers from staircase structures. *Physics of Plasmas*, 24(1):012305, 2017.

[ANB$^+$19] Arash Ashourvan, Raffi Nazikian, E Belli, J Candy, D Eldon, BA Grierson, W Guttenfelder, SR Haskey, C Lasnier, GR McKee, et al. Formation of a high pressure staircase pedestal with suppressed edge localized modes in the diii-d tokamak. *Physical Review Letters*, 123(11):115001, 2019.

[Aro80] Donald G Aronson. Density-dependent interaction–diffusion systems. In *Dynamics and modelling of reactive systems*, pages 161–176. Elsevier, 1980.

[BF71] P.F. Byrd and M.D. Friedman. *Handbook of elliptic integrals for engineers and scientists*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, 1971.

[BI11] Nikolaos A Bakas and Petros J Ioannou. Structural stability theory of two-dimensional fluid flow under stochastic forcing. *Journal of Fluid Mechanics*, 682:332–361, 2011.

[BPH$^+$11] M. Barnes, F. I. Parra, E. G. Highcock, A. A. Schekochihin, S. C. Cowley, and C. M. Roach. Turbulent transport in tokamak plasmas with rotational shear. *Physical Review Letters*, 106:175004, Apr 2011.

[BSM$^+$15] D. Barkley, B. Song, V. Mukund, G. Lemoult, M. Avila, and B. Hof. The rise of fully turbulent flow. *Nature*, 526:550–553, Oct 2015.

[BSY98] NJ Balmforth, Stefan G Llewellyn Smith, and WR Young. Dynamics of interfaces and layers in a stratified turbulent fluid. *Journal of Fluid Mechanics*, 355:329–358, 1998.

[BTW87] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality: An explanation of the 1/f noise. *Physical Review Letters*, 59:381–384, Jul 1987.

[BVO$^+$20] Keaton J Burns, Geoffrey M Vasil, Jeffrey S Oishi, Daniel Lecoanet, and Benjamin P Brown. Dedalus: A flexible framework for numerical simulations with spectral methods. *Physical Review Research*, 2(2):023068, 2020.

[BW85] D. Biskamp and M. Walter. Suppression of shear damping in drift wave turbulence. *Physics Letters A*, 109(1–2):34–38, 1985.

[CD20] Chang-Chun Chen and Patrick H Diamond. Potential vorticity mixing in a tangled magnetic field. *The Astrophysical Journal*, 892(1):24, 2020.

[CH58] John W. Cahn and John E. Hilliard. Free energy of a nonuniform system. i. interfacial free energy. *The Journal of Chemical Physics*, 28(2):258–267, 1958.

[Cha71] Jule G. Charney. Geostrophic turbulence. *Journal of the Atmospheric Sciences*, 28(6):1087–1095, 1971.

[Cha20] Paul Charbonneau. Dynamo models of the solar cycle. *Living Reviews in Solar Physics*, 17(1):1–104, 2020.

[CUH15] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.

[CV91] Fausto Cattaneo and Samuel I Vainshtein. Suppression of turbulent transport by a weak magnetic field. *The Astrophysical Journal*, 376:L21–L24, 1991.

[DBB$^+$00] Andris M Dimits, G Bateman, MA Beer, BI Cohen, W Dorland, GW Hammett, C Kim, JE Kinsey, M Kotschenreuther, AH Kritz, et al. Comparisons and physics basis of tokamak transport models and turbulence simulations. *Physics of Plasmas*, 7(3):969–983, 2000.

[DH93] W. Dorland and G. W. Hammett. Gyrofluid turbulence models with kinetic effects. *Physics of Fluids B: Plasma Physics*, 5(3):812–835, 1993.

[DII10] Patrick H Diamond, Sanae-I Itoh, and Kimitaka Itoh. *Modern Plasma Physics, Volume 1: Physical Kinetics of Turbulent Plasmas*. Cambridge University Press, 2010.

[DIIH05a] Patrick H Diamond, SI Itoh, K Itoh, and TS Hahm. Zonal flows in plasma—a review. *Plasma Physics and Controlled Fusion*, 47(5):R35, 2005.

[DIIH05b] Patrick H Diamond, SI Itoh, K Itoh, and TS Hahm. Zonal flows in plasma—a review. *Plasma Physics and Controlled Fusion*, 47(5):R35, 2005.

[DIIS07] Patrick H Diamond, Sanae-I Itoh, Kimitaka Itoh, and Lara J Silvers. β-plane mhd turbulence and dissipation in the solar tachocline. In *The solar tachocline*, pages 211–239. Cambridge University Press, 2007.

[DLCT94] PH Diamond, Y-M Liang, BA Carreras, and PW Terry. Self-regulating shear flow turbulence: A paradigm for the l to h transition. *Physical review letters*, 72(16):2565, 1994.

[DPDG$^+$10] G. Dif-Pradalier, P. H. Diamond, V. Grandgirard, Y. Sarazin, J. Abiteboul, X. Garbet, Ph. Ghendrih, A. Strugarek, S. Ku, and C. S. Chang. On the validity of the local diffusive paradigm in turbulent plasma transport. *Physical Review E*, 82:025401, Aug 2010.

[DPHG$^+$15] G. Dif-Pradalier, G. Hornung, Ph. Ghendrih, Y. Sarazin, F. Clairet, L. Vermare, P. H. Diamond, J. Abiteboul, T. Cartier-Michaud, C. Ehrlacher, D. Estève, X. Garbet, V. Grandgirard, Ö. D. Gürcan, P. Hennequin, Y. Kosuga, G. Latu, P. Maget, P. Morel, C. Norscini, R. Sabot, and A. Storelli. Finding the elusive $\mathbf{E}\times\mathbf{B}$ staircase in magnetized plasmas. *Physical Review Letters*, 114:085004, Feb 2015.

[DUX$^+$09] BD Dudson, MV Umansky, XQ Xu, PB Snyder, and HR Wilson. BOUT++: A framework for parallel plasma fluid simulations. *Computer Physics Communications*, 180(9):1467–1480, 2009.

[DZB95] J. F. Drake, A. Zeiler, and D. Biskamp. Nonlinear self-sustained drift-wave turbulence. *Physical Review Letters*, 75:4222–4225, Dec 1995.

[FDC19] Xiang Fan, PH Diamond, and L Chacón. Spontaneous transport barriers quench turbulent resistivity in two-dimensional magnetohydrodynamics. *Physical Review E*, 99(4):041201, 2019.

[Fit61] Richard FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 1(6):445–466, 1961.

[GD94] AV Gruzinov and PH Diamond. Self-consistent theory of mean-field electrodynamics. *Physical review letters*, 72(11):1651, 1994.

[GD96] AV Gruzinov and PH Diamond. Nonlinear mean field electrodynamics of turbulent dynamos. *Physics of Plasmas*, 3(5):1853–1857, 1996.

[GD05] ÖD Gürcan and PH Diamond. Dynamics of turbulence spreading in magnetically confined plasmas. *Physics of Plasmas*, 12(3):032303, 2005.

[GD17] Z. B. Guo and P. H. Diamond. Bistable dynamics of turbulence spreading in a corrugated temperature profile. *Physics of Plasmas*, 24(10):100705, 2017.

[GDH$^+$19] Weixin Guo, Patrick H Diamond, David W Hughes, Lu Wang, and Arash Ashourvan. Scale selection and feedback loops for patterns in drift wave-zonal flow turbulence. *Plasma Physics and Controlled Fusion*, 61(10):105002, aug 2019.

[GLB65] P. Goldreich and D. Lynden-Bell. II. Spiral Arms as Sheared Gravitational Instabilities. *Monthly Notices of the Royal Astronomical Society*, 130(2):125–158, 08 1965.

[GLP83] R Grappin, J Leorat, and A Pouquet. Dependence of mhd turbulence spectra on the velocity field-magnetic field correlation. *Astronomy and Astrophysics*, 126:51–58, 1983.

[GM98] DO Gough and ME McIntyre. Inevitability of a magnetic field in the sun's radiative interior. *Nature*, 394(6695):755–757, 1998.

[GS96] L. Gil and D. Sornette. Landau-Ginzburg theory of self-organized criticality. *Physical Review Letters*, 76:3991–3994, May 1996.

[HD18] TS Hahm and PH Diamond. Mesoscopic transport events and the breakdown of fick's law for turbulent fluxes. *Journal of the Korean Physical Society*, 73(6):747–792, 2018.

[HD20] R. A. Heinonen and P. H. Diamond. Turbulence model reduction by deep learning. *Phys. Rev. E*, 101:061201, Jun 2020.

[HDL$^+$04] T S Hahm, P H Diamond, Z Lin, K Itoh, and S-I Itoh. Turbulence spreading into the linearly stable zone and transport scaling. *Plasma Physics and Controlled Fusion*, 46(5A):A323, 2004.

[HM77] Akira Hasegawa and Kunioki Mima. Stationary spectrum of strong turbulence in magnetized nonuniform plasma. *Phys. Rev. Lett.*, 39:205–208, Jul 1977.

[Hor91] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.

[Hor99] W. Horton. Drift waves and transport. *Reviews of Modern Physics*, 71:735–778, Apr 1999.

[HW83] Akira Hasegawa and Masahiro Wakatani. Plasma edge turbulence. *Physical Review Letters*, 50(9):682, 1983.

[IS15] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[ISS$^+$15] K. Ida, Z. Shi, H.J. Sun, S. Inagaki, K. Kamiya, J.E. Rice, N. Tamura, P.H. Diamond, G. Dif-Pradalier, X.L. Zou, et al. Towards an emerging understanding of non-locality phenomena and non-local transport. *Nuclear Fusion*, 55(1):013022, 2015.

[ITT$^+$13] S. Inagaki, T. Tokuzawa, N. Tamura, S.-I. Itoh, T. Kobayashi, K. Ida, T. Shimozuma, S. Kubo, K. Tanaka, T. Ido, et al. How is turbulence intensity determined by macroscopic variables in a toroidal plasma? *Nuclear Fusion*, 53(11):113006, 2013.

[Jay15] Laurent O. Jay. Lobatto methods. In Björn Engquist, editor, *Encyclopedia of Applied and Computational Mathematics*, pages 817–826. Springer-Verlag, Berlin, 2015.

[Kad65] B. B. Kadomtsev. *Plasma Turbulence*. Academic Press, 1965.

[KB14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[KD03] Eun-jin Kim and P. H. Diamond. Zonal flows and transient dynamics of the $l-h$ transition. *Physical Review Letters*, 90:185006, May 2003.

[KDM$^+$03] E. Kim, P.H. Diamond, M. Malkov, T.S. Hahm, K. Itoh, S.-I. Itoh, S. Champeaux, I. Gruzinov, O. Gurcan, and C. Holland. Non-perturbative models of intermittency in drift-wave turbulence: towards a probabilistic theory of anomalous transport. *Nuclear Fusion*, 43(9):961, 2003.

[KIO91] George Em Karniadakis, Moshe Israeli, and Steven A Orszag. High-order splitting methods for the incompressible Navier-Stokes equations. *Journal of computational physics*, 97(2):414–443, 1991.

[Kra67] Robert H Kraichnan. Inertial ranges in two-dimensional turbulence. *The Physics of Fluids*, 10(7):1417–1423, 1967.

[Law57] John D Lawson. Some criteria for a power producing thermonuclear reactor. *Proceedings of the physical society. Section B*, 70(1):6, 1957.

[LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[LH04] Z. Lin and T. S. Hahm. Turbulence spreading and transport scaling in global gyrokinetic particle simulations. *Physics of Plasmas*, 11(3):1099–1108, 2004.

[LLPS93] Moshe Leshno, Vladimir Ya Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, 1993.

[LPS14] Gabriel J Lord, Catherine E Powell, and Tony Shardlow. *An introduction to computational stochastic PDEs*, volume 50. Cambridge University Press, 2014.

[LPW$^+$17] Zhou Lu, Hongming Pu, Feicheng Wang, Zhiqiang Hu, and Liwei Wang. The expressive power of neural networks: A view from the width. In *Advances in neural information processing systems*, pages 6231–6239, 2017.

[MBW+19] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre G.R. Day, Clint Richardson, Charles K. Fisher, and David J. Schwab. A high-bias, low-variance introduction to machine learning for physicists. *Physics Reports*, 810:1 – 124, 2019.

[Mof78] Henry K Moffatt. *Magnetic field generation in electrically conducting fluids*. Cambridge University Press, Cambridge-London-New York-Melbourne, 1978.

[Mor98] O.A. Mornev. Modification of the Biot method on the basis of the principal of minimum dissipation (with an application to the problem of propagation of nonlinear concentration waves in an autocatalytic medium). *Zhurnal Fizicheskoi Khimii*, 72(1):124–131, 1998.

[Mur02] James D. Murray. *Mathematical Biology I. An Introduction*. Springer, New York, 3 edition, 2002.

[NAY62] J. Nagumo, S. Arimoto, and S. Yoshizawa. An active pulse transmission line simulating nerve axon. *Proceedings of the IRE*, 50(10):2061–2070, 1962.

[NBD07] Ryusuke Numata, Rowena Ball, and Robert L Dewar. Bifurcation in electrostatic resistive drift wave turbulence. *Physics of Plasmas*, 14(10):102312, 2007.

[New80] William I Newman. Some exact solutions to a non-linear diffusion problem in population genetics and combustion. *Journal of Theoretical Biology*, 85(2):325–334, 1980.

[NNR05] V. Naulin, A. H. Nielsen, and J. Juul Rasmussen. Turbulence spreading, anomalous transport, and pinch effect. *Physics of Plasmas*, 12(12):122306, 2005.

[NSK+05] R Nazikian, K Shinohara, GJ Kramer, E Valeo, K Hill, TS Hahm, G Rewoldt, S Ide, Y Koide, Y Oyama, et al. Measurement of turbulence decorrelation during transport barrier evolution in a high-temperature fusion plasma. *Physical Review Letters*, 94(13):135002, 2005.

[Par15] Jeffrey B Parker. *Zonal flows and turbulence in fluids and plasmas*. PhD thesis, Princeton University, 2015.

[Ped05] Morten Gram Pedersen. Wave speeds of density dependent Nagumo diffusion equations – inspired by oscillating gap-junction conductance in the islets of Langerhans. *Journal of Mathematical Biology*, 50(6):683–698, 2005.

[PM80] Yves Pomeau and Paul Manneville. Intermittent transition to turbulence in dissipative dynamical systems. *Comm. Math. Phys.*, 74(2):189–197, 1980.

[Pom86] Y. Pomeau. Front motion, metastability and subcritical bifurcations in hydrodynamics. *Physica D: Nonlinear Phenomena*, 23(1):3 – 11, 1986.

[Pom15] Y. Pomeau. The transition to turbulence in parallel flows: A personal view. *Comptes Rendus Mecanique*, 343:210–218, March 2015.

[Pou78] A Pouquet. On two-dimensional magnetohydrodynamic turbulence. *Journal of Fluid Mechanics*, 88(1):1–16, 1978.

[Pra25] Ludwig Prandtl. Bericht über untersuchungen zur ausgebildeten turbulenz. *Z. Angew. Math. Meth.*, 5:136–39, 1925.

[Rhi75] Peter B Rhines. Waves and turbulence on a beta-plane. *Journal of Fluid Mechanics*, 69(3):417–443, 1975.

[RPSD16] DE Ruiz, JB Parker, EL Shi, and IY Dodin. Zonal-flow dynamics from a phase-space perspective. *Physics of Plasmas*, 23(12):122304, 2016.

[RVBS17] David Rolnick, Andreas Veit, Serge Belongie, and Nir Shavit. Deep learning is robust to massive label noise. *arXiv preprint arXiv:1705.10694*, 2017.

[Sco90] Bruce D. Scott. Self-sustained collisional drift-wave turbulence in a sheared magnetic field. *Physical Review Letters*, 65:3289–3292, Dec 1990.

[Sco92] Bruce D. Scott. The mechanism of self-sustainment in collisional drift wave turbulence. *Physics of Fluids B: Plasma Physics*, 4(8):2468–2494, 1992.

[SD99] AI Smolyakov and PH Diamond. Generalized action invariants for drift waves-zonal flow systems. *Physics of Plasmas*, 6(12):4410–4413, 1999.

[SD20] Rameswar Singh and P. H. Diamond. When does turbulence spreading matter? *Physics of Plasmas*, 27(4):042308, 2020.

[SG69] R.Z̃. Sagdeev and A.Ã. Galeev. *Nonlinear Plasma Theory*. Benjamin, New York ad London, 1969.

[SG98] Y. Sarazin and Ph. Ghendrih. Intermittent particle transport in two-dimensional edge turbulence. *Physics of Plasmas*, 5(12):4214–4228, 1998.

[SGM97] Faustino Sánchez-Garduño and Philip K. Maini. Travelling wave phenomena in non-linear diffusion degenerate Nagumo equations. *Journal of Mathematical Biology*, 35(6):713–728, Jun 1997.

[SII$^+$12] S. Sugita, K. Itoh, S.-I. Itoh, M. Yagi, G. Fuhr, P. Beyer, and S. Benkadda. Ballistic propagation of turbulence front in tokamak edge plasmas. *Plasma Physics and Controlled Fusion*, 54(12):125001, 2012.

[SY12] Kaushik Srinivasan and WR Young. Zonostrophic instability. *Journal of the atmospheric sciences*, 69(5):1633–1656, 2012.

[SZ92] EA Spiegel and J-P Zahn. The solar tachocline. *Astronomy and Astrophysics*, 265:106–114, 1992.

[TDH07] Steven M Tobias, Patrick H Diamond, and David W Hughes. β-plane magnetohydrodynamic turbulence in the solar tachocline. *The Astrophysical Journal Letters*, 667(1):L113, 2007.

[Ter00] P. W. Terry. Suppression of turbulence and transport by sheared flow. *Rev. Mod. Phys.*, 72:109–165, Jan 2000.

[TNG13] Natalia Tronko, Sergey V Nazarenko, and Sebastien Galtier. Weak turbulence in two-dimensional magnetohydrodynamics. *Physical Review E*, 87(3):033103, 2013.

[TS15] Geoffrey Ingram Taylor and William Napier Shaw. I. eddy motion in the atmosphere. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 215(523-537):1–26, 1915.

[VCMMS15] B. Van Compernolle, G. J. Morales, J. E. Maggs, and R. D. Sydora. Laboratory study of avalanches in magnetized plasmas. *Physical Review E*, 91:031102, Mar 2015.

[VVS61] AA Vedenov, EP Velikhov, and RZ Sagdeev. Nonlinear oscillations of rarified plasma. *Nuclear Fusion*, 1(2):82, 1961.

[VVS62] A.A. Vedenov, E.P. Velikhov, and R.Z. Sagdeev. Quasi-linear theory of plasma oscillations. *Nucl. Fusion, Suppl.*, 2:465, 1962.

[vWHS$^+$16] F. van Wyk, E. G. Highcock, A. A. Schekochihin, C. M. Roach, A. R. Field, and W. Dorland. Transition to subcritical turbulence in a tokamak plasma. *Journal of Plasma Physics*, 82(6):905820609, 2016.

[WABC06] R. E. Waltz, M. E. Austin, K. H. Burrell, and J. Candy. Gyrokinetic simulations of off-axis minimum-q profile corrugations. *Physics of Plasmas*, 13(5):052301, 2006.

[WBB$^+$82] Fritz Wagner, G Becker, K Behringer, D Campbell, A Eberhagen, W Engelhardt, G Fussmann, O Gehre, J Gernhardt, G v Gierke, et al. Regime of improved confinement and high beta in neutral-beam-heated divertor discharges of the asdex tokamak. *Physical Review Letters*, 49(19):1408, 1982.

[WH84] Masahiro Wakatani and Akira Hasegawa. A collisional drift wave description of plasma edge turbulence. *The Physics of Fluids*, 27(3):611–618, 1984.

[YKDH14] S. Yi, J. M. Kwon, P. H. Diamond, and T. S. Hahm. Effects of q-profile structure on turbulence spreading: A fluctuation intensity transport analysis. *Physics of Plasmas*, 21(9):092509, 2014.

[Zel57] Ya B Zeldovich. The magnetic field in the two-dimensional motion of a conducting turbulent fluid. *Sov. Phys. JETP*, 4:460–462, 1957.

[ZLF12] Vladimir E Zakharov, Victor S L'vov, and Gregory Falkovich. *Kolmogorov spectra of turbulence I: Wave turbulence*. Springer Science & Business Media, 2012.