Title

Learning how structures form in drift-wave turbulence

Journal

Plasma Physics and Controlled Fusion, 62(10)

ISSN

0741-3335

Authors

Heinonen, RA
Diamond, PH

Publication Date

2020-10-01

DOI

10.1088/1361-6587/abad02

Learning how structures form in drift-wave turbulence

Learning how structures form in drift-wave turbulence

R. A. Heinonen[1] and P. H. Diamond[1]

University of California San Diego, La Jolla, California 92093

(Dated: 3 August 2020)

Drift-wave turbulence produces anomalous transport via cross-correlations between fluctuations. This transport has profound implications for confinement, structure formation, and virtually all aspects of the nonlinear turbulent dynamics. In this work, we use a data-driven method based on deep learning in order to study turbulent transport in the 2-D Hasegawa-Wakatani system and infer a reduced mean-field model from numerical solution. In addition to the usual turbulent diffusion, we find an effect which couples the particle flux to the local *gradient* of vorticity, which tends to modulate the density profile. The direct coupling to the shear is relatively weak. In addition, the deep learning method finds a model for spontaneous zonal flow generation by negative viscosity, stabilized by nonlinear and hyperviscous terms. We compare these results to analytic calculations using quasilinear theory and wave kinetics, finding qualitative agreement, though the calculations miss certain higher-order effects. A simplified, 1-D model for the evolution of the profile, flow, and intensity based on the deep learning results is solved numerically and compared to previous models for staircasing based on bistability. We see that the physics uncovered by the deep learning method provided simple explanations for the formation of zonal structures in the density, flow, and turbulence fields. We highlight the important role of symmetry in the deep learning method and speculate on the portability of the method to other applications.

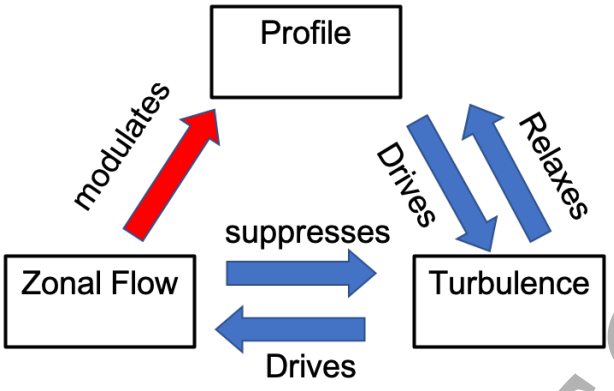


FIG. 1: Cartoon depicting the basic feedback loops in the drift-wave/ZF system. The interaction between ZF and profile is especially pertinent to this work and is highlighted in red.

I. INTRODUCTION

Drift-wave turbulence[1] 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)[2], 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[3]. In addition, features such as quasiperiodic staircases[4–6] 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 turbulence induces a flux which relaxes the driving profile, tending to quench the instability (see Fig. 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[7]). 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[8]

$$\Gamma \simeq -\frac{\gamma_{\mathbf{k}}}{k_\perp^2}\frac{d\langle n\rangle}{dx}, \tag{1}$$

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

$$\mathscr{M}_\xi : (\langle\psi_1\rangle,\ldots,\langle\psi_n\rangle)|_{r_0,t_0} \mapsto \langle\tilde{v}_r(r_0,t_0)\tilde{\xi}(r_0,t_0)\rangle \tag{2}$$

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[9,10]

$$\partial_t n + \{\phi, n\} = C(n-\phi) \tag{3}$$

$$\partial_t \nabla_\perp^2\phi + \{\phi, \nabla_\perp^2\phi\} = C(n-\phi). \tag{4}$$

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[11,12] (the adiabatic limit of HW). It is useful to separate the mean and fluctuating parts of Eqs. (3–4) to obtain

$$\partial_t \tilde{n} + N'\partial_y\tilde{\phi} + V_y\partial_y\tilde{n} = C(\tilde{n} - \tilde{\phi}) \tag{5}$$

$$\partial_t \tilde{\nabla}_\perp^2\phi - V_y''\partial_y\tilde{\phi} + V_y\partial_y\nabla_\perp^2\tilde{\phi} = C(\tilde{n} - \tilde{\phi}) \tag{6}$$

$$\partial_t N + \partial_x\Gamma = 0 \tag{7}$$

$$\partial_t V_y' - \partial_x^2\Pi = 0, \tag{8}$$

where we have used the Taylor identity[13] 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[14,15] 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. IV E.) 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[16] 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*[17–19], 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[20].

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), 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 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.[21]. 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

[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.

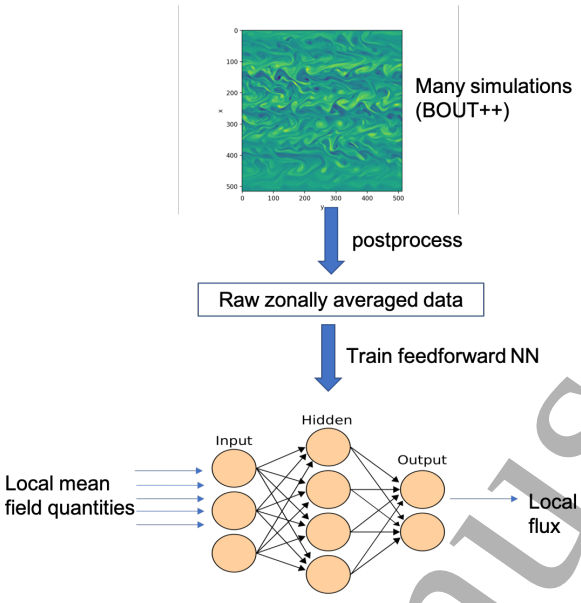


FIG. 2: Basic schematic of the deep learning method. Simulations are ran, post-processed, and then fed into a feedforward neural network.

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.

II. METHODS

A. Numerical solution of 2-D HW

The workflow of the method is illustrated in Fig. 2. We first perform direct numerical simulation (DNS) of the (modified) HW system[22] on a 2-D slab

$$\partial_t n + N'(x)\partial_y \phi + \{\phi, n\} = \alpha(\tilde{\phi} - \tilde{n}) - D\nabla^4 n \tag{9}$$

$$\partial_t \nabla^2 \phi + \{\phi, \nabla^2 \phi\} = \alpha(\tilde{\phi} - \tilde{n}) - \mu\nabla^2 \phi + D\nabla^6 \phi. \tag{10}$$

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[23]. 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[24] — 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[25], this system is solved on a square 512×512 spatial grid using the Karniadakis time-stepping algorithm[26]. 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 $n = 1, 2, 3$.

B. Post-processing

Next, zonally averaged quantities of interest—namely, the turbulence intensity (here represented by the turbulent PE ε—see Sec. II D), 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, $128N_t = 254720$ data points representing the map (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.

C. 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.[27], 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

$$x_j^{i+1} = \sigma\left(\mathbf{w}_j^i \cdot \mathbf{x}^i + b_j^i\right). \tag{11}$$

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"[28]

$$f(x) = \begin{cases} x, & x \geq 0 \\ e^x - 1, & x < 0, \end{cases} \tag{12}$$

a smoother alternative to the ReLU, as our hidden layer activation function. Batch normalization (BN)[29], 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[30]. 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

$$L = \sum_i \ln\left(\cosh(y_i^* - f_W(\mathbf{x}_i))\right) + \lambda ||W||^2, \tag{13}$$

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.

D. 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). 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

$$\begin{cases} \phi \to \phi + v_0 x \\ y \to y - v_0 t. \end{cases} \tag{14}$$

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 stabiliz-

ing 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 $\varepsilon = \langle(\tilde{n} - \nabla_\perp \tilde{\phi})^2\rangle$[14,15]. 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

$$x \to -x, y \to -y; \tag{15}$$

$$x \to -x, \phi \to -\phi, n \to -n; \tag{16}$$

$$y \to -y, \phi \to -\phi, n \to -n. \tag{17}$$

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

$$\tilde{\Gamma}(\varepsilon, N', U, U', U'') = \frac{1}{4}(\Gamma(\varepsilon, N', U, U', U'') + \Gamma(\varepsilon, N', -U, U', -U'') \tag{18}$$

$$-\Gamma(\varepsilon, -N', U, -U', -U'') - \Gamma(\varepsilon, -N', -U, -U', -U'')). \tag{19}$$

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

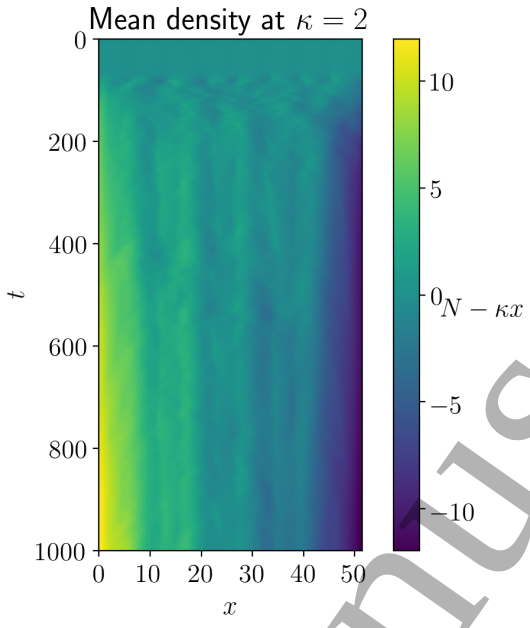


FIG. 3: Plot of change in mean density $N - \kappa x$, where κ is the initial gradient drive, at $\kappa = 2$, from BOUT++ simulation.

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 which we have checked against in this work are entropy production[31] 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.

III. RESULTS

A. Numerical solution

In Figs. 3–5 we show the mean density, vorticity, and turbulent PE from a typical numerical solution of Eqs. (9), with gradient drive $\kappa = 2$. A few characteristics of the self-organization

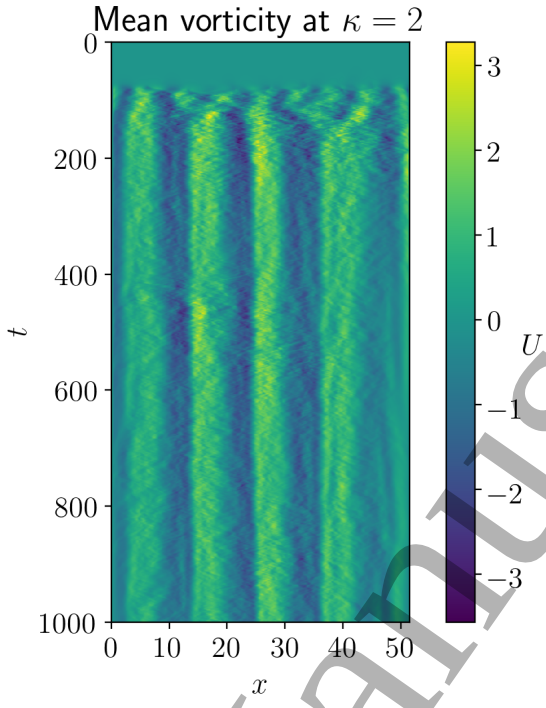


FIG. 4: Plot of mean vorticity U at $\kappa = 2$, from BOUT++ simulation.

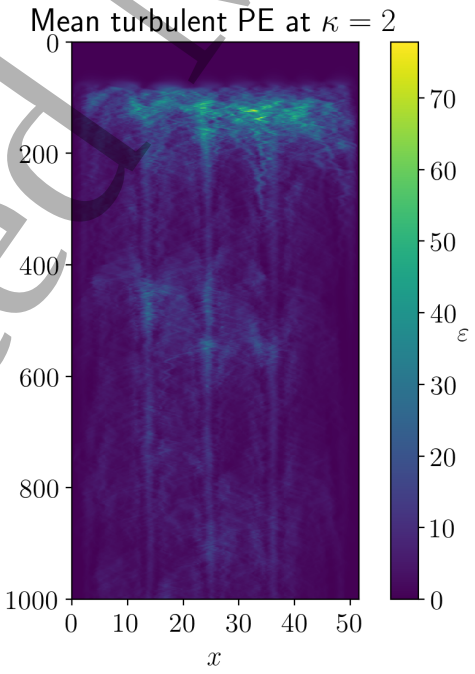
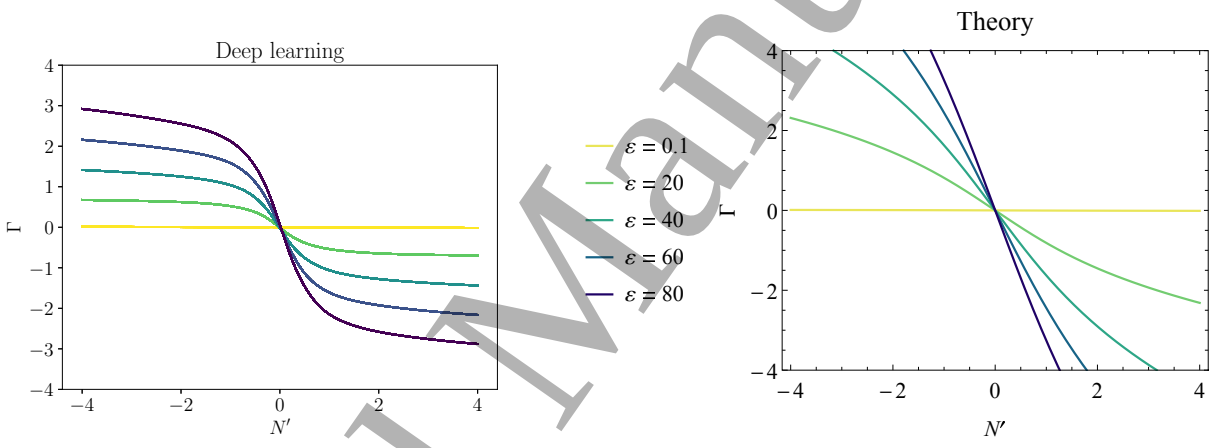


FIG. 5: Plot of mean turbulent PE ε at $\kappa = 2$, from BOUT++ simulation. Note the appearance of corrugations near $x = 13, 25, 37$.

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, 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.

B. Particle flux

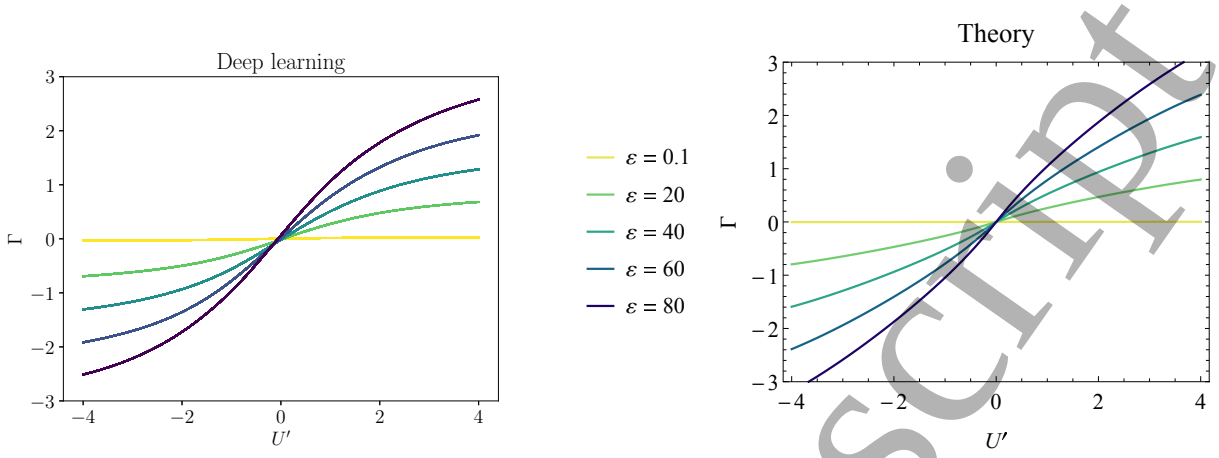


(a) (b)

FIG. 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)[21]) shows the behavior learned by the deep neural network, whereas the righthand figure shows the prediction of the simple analytical model from Sec. IV using the ansatz Lorentzian spectrum Eq. 36. The analytical model shows reasonable agreement when N' is small ($\lesssim 1$), but fails to accurately model saturation effects.

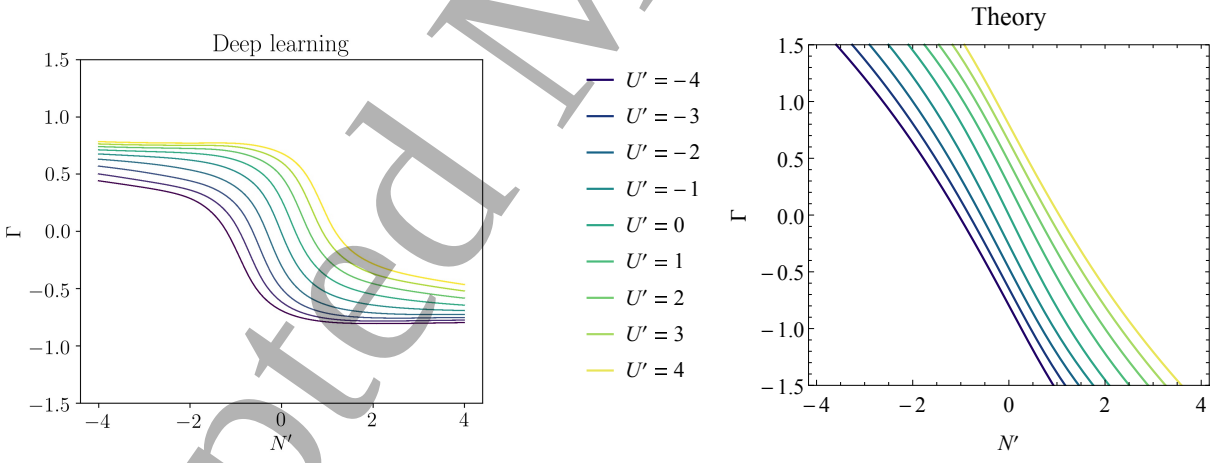
The DNN finds that the particle flux depends most strongly on N', U', and ε. The flux does *not* noticeably depend on U''. The basic leading-order behavior can be summarized as

$$\Gamma \simeq \varepsilon(-D_n N' + D_u U'), \tag{20}$$



(a) (b)

FIG. 7: Same as Fig. 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)[21].



(a) (b)

FIG. 8: Same as Figs. 6–7, 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)[21].

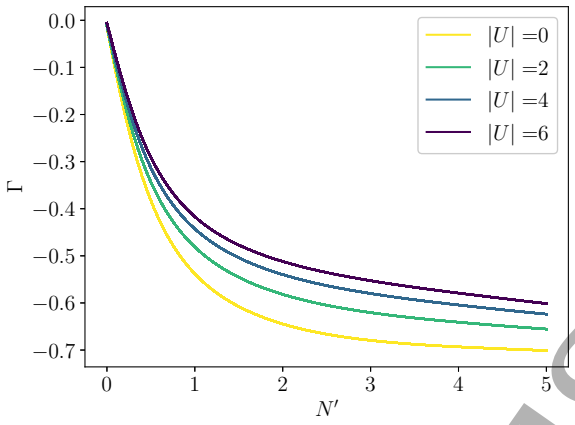


FIG. 9: Learned particle flux at fixed $U' = U'' = 0$ and fixed $\varepsilon = 20$, as a function of N' and $|U|$ (the curves with $\pm U$ lie on top of each other). The flux is reduced by an approximate factor $(1 + 0.04|U|)^{-1}$.

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, 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.[21]. 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

$$\partial_t N = -\partial_x \Gamma = D_u \varepsilon_0 U_0 q^2 \sin qx + \dots \tag{21}$$

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 they become large, and can no longer be simply expressed as the simple linear combination (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.

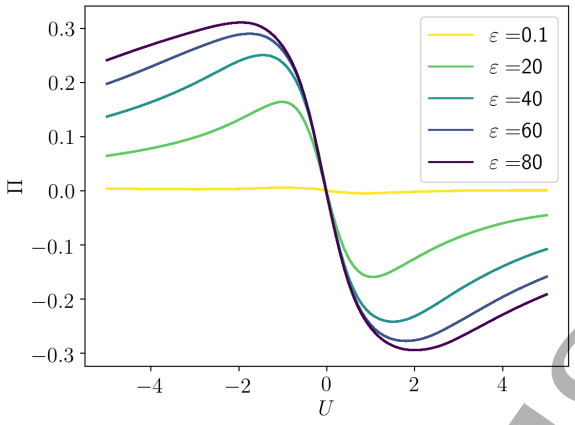


FIG. 10: Plot of learned Reynolds stress against vorticity U at fixed $N' = 2$ and $U' = U'' = 0$ and several values of the intensity. The basic behavior is that of a negative viscosity, stabilized by nonlinear effects. From Heinonen and Diamond (2020)[21].

In a final effect, which is missed by the analytical calculation in Sec. IV, 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[32], 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. 6–9.

C. Reynolds stress

The Reynolds stress model learned by the DNN depends strongly on U and ε, as shown in Fig. 10. At small U, the leading behavior is of the form

$$\Pi = \varepsilon(-\chi_1 U + \chi_3 U^3) \tag{22}$$

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. 11), which is typical of strong turbulence scaling[1].

The DNN also detects a stabilizing turbulent hyperdiffusion term roughly of the form $-\varepsilon\chi_4 U''$,

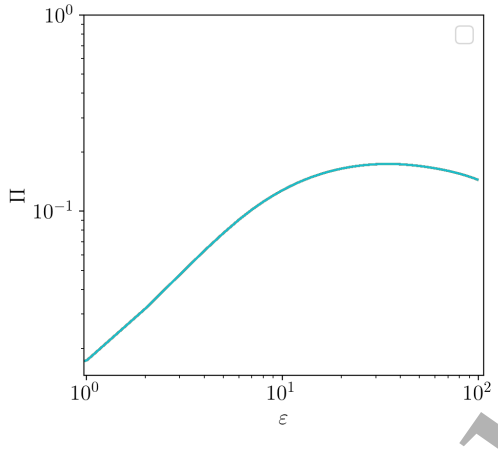


FIG. 11: Log-log plot of learned Reynolds stress against intensity at fixed $U = 0.5$, $N' = 2$, $U' = U'' = 0$ and several values of the intensity. The scaling exponent is unity for low to moderate values of the intensity but decreases as ε increases.

.

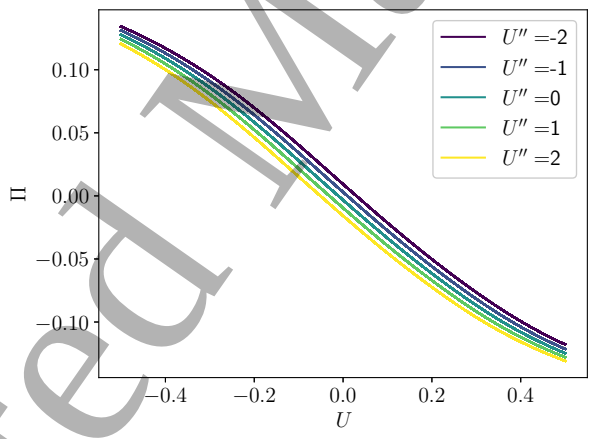


FIG. 12: Plot of learned Reynolds stress against vorticity U at fixed $N' = 2$, $U' = 0$, $\varepsilon = 20$, and several values of U''. The leading order contribution from U'' is a stabilizing linear term. From Heinonen and Diamond (2020)[21].

.

with $\chi_4 \sim 0.0005$ (Figs. 12–13). 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.

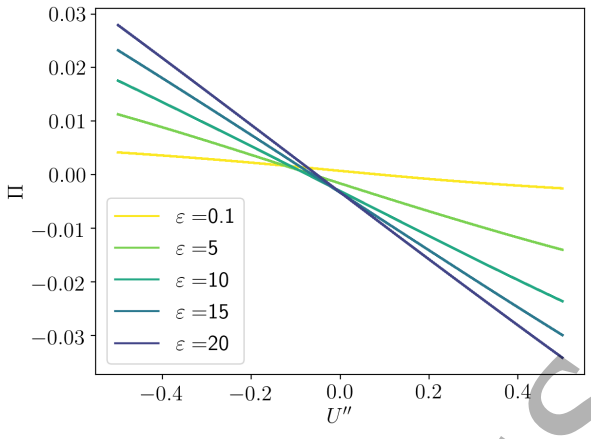


FIG. 13: Plot of learned Reynolds stress against U'' at fixed $N' = 2$, $U = U' = 0$ and several values of the intensity. We should have $\Pi \to -\Pi$ under $U'' \to -U''$ here, but the model fails to precisely learn this, which may be attributed to the relatively small contribution to the loss function from the hyperdiffusion term. However, it is clear that this term scales roughly linearly with intensity. From Heinonen and Diamond (2020)[21].

.

The above findings are consistent with higher-order quasilinear theory, from which one can obtain the model equation[2]

$$\partial_t U = \partial_x^2(-D_1 U + D_3 U^3 - D_4 U''). \tag{23}$$

This equation, which might also be anticipated using Ginzburg-Landau theory, has the form of a 1-D Cahn-Hilliard equation[33] 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[34]. 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[35,36].

The Reynolds stress is also found to be moderately reduced in the presence of a nonzero N' or

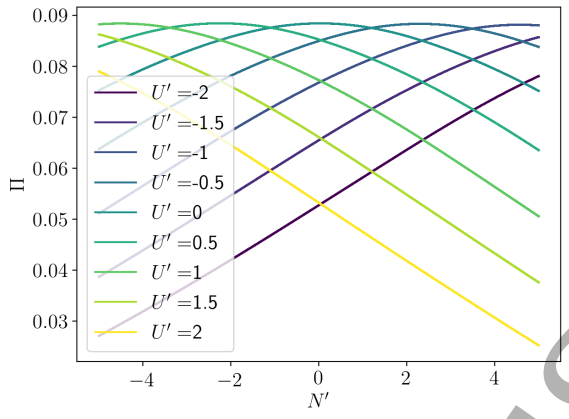


FIG. 14: Plot of learned Reynolds stress against N' at fixed $U = 1$,$\varepsilon = 20$, $U'' = 0$, and several values of U'. The presence of a gradient in U' or N' tends to suppress the Reynolds stress. From Heinonen and Diamond (2020)[21].

.

U', by an overall factor f which behaves roughly as

$$f = \frac{1}{1 + a(N' + bU')^2} \tag{24}$$

with $b \approx 4$ and $a \approx 0.04$. In Fig. 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.

IV. THEORY

A. Linear theory

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

$$\partial_t \tilde{n} + N' \partial_y \tilde{\phi} + V_y \partial_y \tilde{n} = \alpha(\tilde{\phi} - \tilde{n}) \tag{25}$$

$$\partial_t \nabla_\perp^2 \tilde{\phi} - V_y'' \partial_y \tilde{\phi} + V_y \partial_y \nabla_\perp^2 \tilde{\phi} = \alpha(\tilde{\phi} - \tilde{n}). \tag{26}$$

(27)

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

$$(\omega - k_y V_y)^2 - \left(k_y k^{-2} V_y'' - i\alpha(1+k^{-2})\right)(\omega - k_y V) + i\alpha k_y (N' + V_y'') = 0. \tag{28}$$

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

$$\omega_r = k_y V_y + a/2 \pm \sqrt{\frac{1}{8}(a^2 - b^2 + d)} \tag{29}$$

$$\gamma = -b/2 \pm \sqrt{\frac{1}{8}(b^2 - a^2 + d)}, \tag{30}$$

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

$$\omega_r = \frac{k_y (N' + V_y'')}{1+k^2} \tag{31}$$

$$\gamma = \frac{k_y^2}{\alpha(1+k^2)^3}(N' + V_y'')(k^2 N' - V_y''), \tag{32}$$

where we have suppressed the Doppler shift by writing $\mathrm{Re}\,\omega = \omega_r + k_y V_y$. 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.

B. Quasilinear fluxes and wave-kinetic equation

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

$$\Gamma = \mathrm{Re}\int d^2\mathbf{k} - ik_y \tilde{n}_{\mathbf{k}} \tilde{\phi}^*_{\mathbf{k}} \tag{33}$$

$$= \int d^2\mathbf{k} \frac{-k_y^2 N'(\gamma_{\mathbf{k}}+\alpha) + \alpha k_y \omega_{\mathbf{k}}}{\omega_{\mathbf{k}}^2 + (\gamma_{\mathbf{k}}+\alpha)^2} |\tilde{\phi}_{\mathbf{k}}|^2 \tag{34}$$

$$\simeq -\frac{1}{\alpha}\int d^2\mathbf{k} \frac{k_y^2}{1+k^2}\left(k^2 N' - U'\right)|\tilde{\phi}_{\mathbf{k}}|^2 \tag{35}$$

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. 6–8, we plot the analytical result (34) using the exact frequencies (29)–(30) and an ansatz Lorentzian spectrum centered about the most unstable mode

$$\varepsilon_{\mathbf{k}} = \frac{\varepsilon}{2\pi^2 \Delta k_x \Delta k_y} \frac{1}{1+k_x^2/\Delta k_x^2}\left(\frac{1}{1+(k_y-\sqrt{2})^2/\Delta k_y^2} + \frac{1}{1+(k_y+\sqrt{2})^2/\Delta k_y^2}\right), \tag{36}$$

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

$$\Pi = \int d^2\mathbf{k}\, k_x k_y |\tilde{\phi}_{\mathbf{k}}|^2. \tag{37}$$

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[37] to obtain the wave-kinetic equation (WKE)

$$\partial_t \varepsilon_{\mathbf{k}} + \partial_{k_x}\omega_{\mathbf{k}} \partial_x \varepsilon_{\mathbf{k}} - \left(\partial_x \omega_{\mathbf{k}} + k_y U\right)\partial_{k_x}\varepsilon_{\mathbf{k}} = (2\gamma_{\mathbf{k}} + \partial_{k_x,x}^2 \omega_{\mathbf{k}})\varepsilon_{\mathbf{k}}. \tag{38}$$

The unusual term $\partial_{k_x,x}^2 \omega_{\mathbf{k}}$ comes from the non-Hermiticity of the time evolution operator in the fluctuation equations (25–26) and can be derived with the Wigner-Moyal formalism[38]. 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. (38) is consistent with the equation for the mean turbulent PE

$$\partial_t \varepsilon + 2(\Gamma - \partial_x \Pi)(N' + U') = 0. \tag{39}$$

Equation (39) results from subtracting (26) from (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.

C. 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

$$k_x = k_{x0} - k_y U t. \tag{40}$$

Such a use of the shearing coordinates of Goldreich and Lynden-Bell[39] follows, for example, Kim and Diamond (1999)[3]. Taking $k_{x0} = 0$, we then have a solution

$$\varepsilon_{\mathbf{k}}(t) = \varepsilon \exp\left(2\gamma_k|_{k_x = -k_y U t} t\right). \tag{41}$$

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

$$\partial_t U = -\partial_x^2 \left[\varepsilon \int dk_y \frac{k_y^2 U \tau}{(1 + k_y^2(1 + U^2\tau^2))^2} \exp\left(\frac{2k_y^2 \tau}{\alpha(1 + (1 + U^2\tau^2)k_y^2)^3} (N' + U')(k_y^2(1 + U^2\tau^2)N' - U') \right) \right]. \tag{42}$$

The integral (call it I) is plotted for $N' = 2, U' = 0, \alpha = 2, \tau = 0.5$ in Fig. 15. In the limit $U \to \infty$ one has the asymptotic behavior

$$I \simeq \frac{\pi U \tau}{2(1 + U^2\tau^2)^{3/2}} \sim U^{-2}. \tag{43}$$

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

$$I \simeq aU\tau - bU^3\tau^3$$

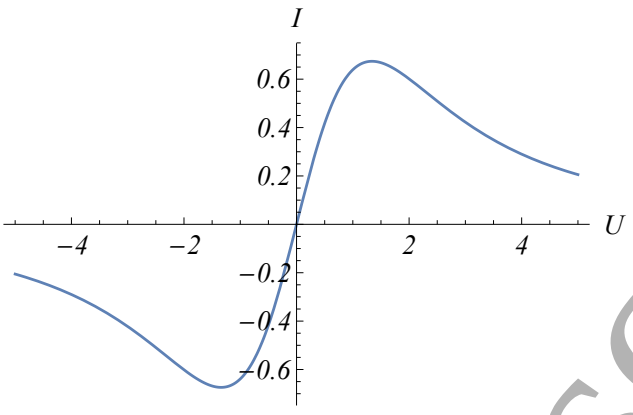


FIG. 15: Plot of integral I appearing in Eq. (42), for $N' = 2, U' = 0, \alpha = 2, \tau = 0.5$. Compared to Fig. 10, the Reynolds stress is overestimated by an order of magnitude.

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

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

and

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

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. 42 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.

D. Reduced 1-D model

The WKE, together with the evolution equations

$$\partial_t N = \partial_x \int d^2\mathbf{k} \frac{k_y^2}{\alpha(1+k^2)^3}(k^2 N' - U')\varepsilon_{\mathbf{k}} \tag{44}$$

$$\partial_t U = \partial_x^2 \int d^2\mathbf{k} \frac{k_x k_y}{(1+k^2)^2}\varepsilon_{\mathbf{k}}, \tag{45}$$

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,

$$\Gamma = \frac{-D_n \varepsilon N' + D_u \varepsilon U'}{1 + c_1 N'^2 + c_2 U'^2} \tag{46}$$

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

$$\Pi = -\frac{c_3 \varepsilon}{1 + a(N' + bU')^2} \frac{\tanh U/U_0}{(1 + c_4 U^2)^{1/2}} - \chi_4 \varepsilon U'', \tag{47}$$

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. (39), whence we have the system

$$\partial_t N = \partial_x \left(\frac{D_n \varepsilon \partial_x N - D_u \varepsilon \partial_x U}{1 + c_1 (\partial_x N)^2 + c_2 (\partial_x U)^2} \right) - D_0 \partial_x^4 N \tag{48}$$

$$\partial_t U = \partial_x^2 \left(-\frac{c_3 \varepsilon}{1 + a(N' + bU')^2} \frac{\tanh U/U_0}{(1 + c_4 U^2)^{1/2}} - \chi_4 \varepsilon \partial_x^2 U \right) - \mu U - D_0 \partial_x^4 U \tag{49}$$

$$\partial_t \varepsilon = \left[\frac{D_n \varepsilon \partial_x N - D_u \varepsilon \partial_x U}{1 + c_1 (\partial_x N)^2 + c_2 (\partial_x U)^2} + \partial_x \left(-\frac{c_3 \varepsilon}{1 + a(N' + bU')^2} \frac{\tanh U/U_0}{(1 + c_4 U^2)^{1/2}} - \chi_4 \varepsilon \partial_x^2 U \right) \right] (\partial_x N + \partial_x U) \tag{50}$$

$$- \gamma_d \varepsilon - \gamma_{NL} \varepsilon^2 + D_\varepsilon \partial_x^2 \varepsilon. \tag{51}$$

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. (28). 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

$$N(x,t=0) = \kappa x, \tag{52}$$

U with a small inhomogeneity

$$U(x,t=0) = 0.001 \sin\frac{6\pi x}{L}, \tag{53}$$

and ε with a small uniform intensity

$$\varepsilon(x,t=0) = \varepsilon_0 \tag{54}$$

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

$$N'(x=0,t) = N'(x=L,t) = \kappa, \tag{55}$$

$$N(x=0,t) = 0, \tag{56}$$

$$N(x=L,t) = \kappa L, \tag{57}$$

$$U(x=0,t) = U''(x=0,t) \tag{58}$$

$$= U(x=L,t) = U''(x=L,t) = 0, \tag{59}$$

$$\varepsilon(x=0,t) = \varepsilon(x=L,t) = \varepsilon_0. \tag{60}$$

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[40] with grid spacing $\Delta x = 0.01$.

The solutions are shown in Figs. 16–18. We have checked that they properly conserve U, N, and the total PE to good approximation.

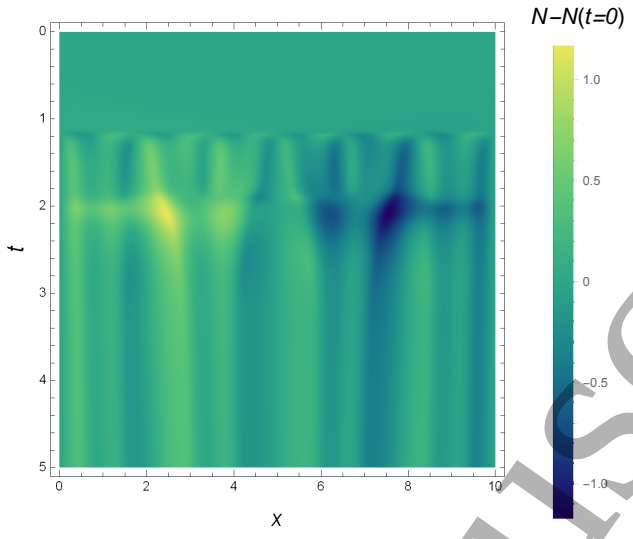


FIG. 16: Color map showing solution $N - N(t = 0)$ to 1-D model with gradient drive $\kappa = 1.5$.

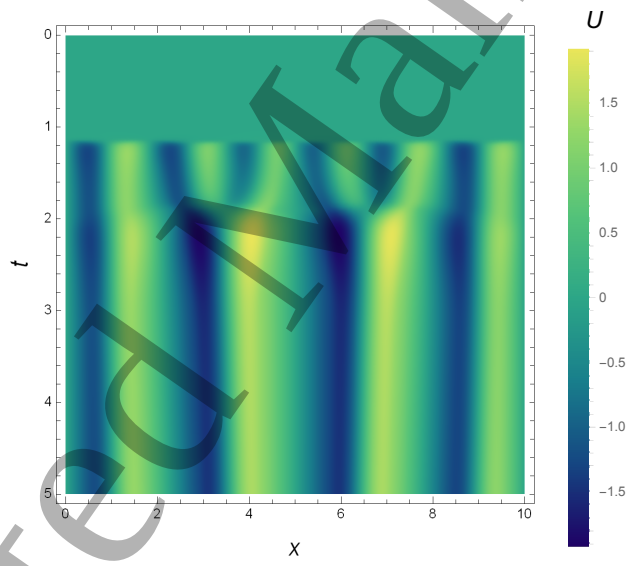


FIG. 17: Color map showing solution U to 1-D model with gradient drive $\kappa = 1.5$.

There are three basic stages of evolution, illustrated by time slices in Figs. 19–21. 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 occurs

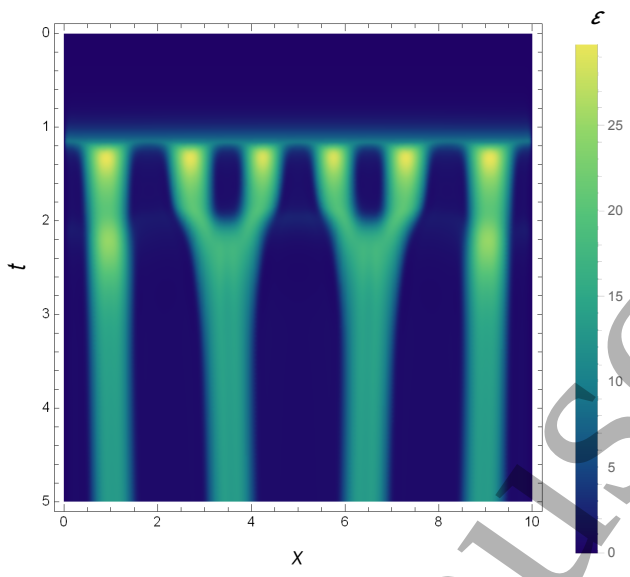


FIG. 18: Color map showing solution ε to 1-D model with gradient drive $\kappa = 1.5$.

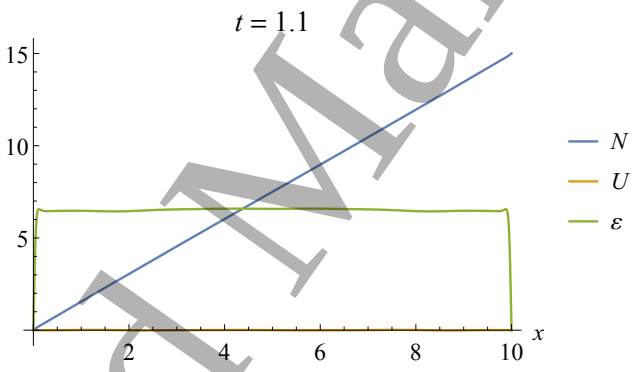


FIG. 19: Plot of solution to 1-D model at $t = 1.1$, illustrating first stage of evolution. The turbulence field grows uniformly with the density gradient as the free energy source. No zonal flow has developed.

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

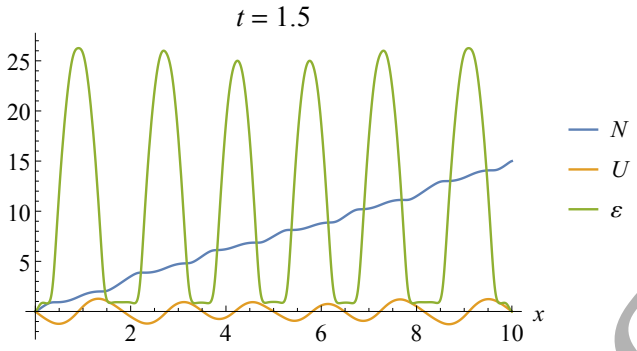


FIG. 20: Plot of solution to 1-D model at $t = 1.5$, illustrating second stage of evolution. A zonal flow spontaneously forms, corrugating the intensity profile and inducing a staircase in the density profile.

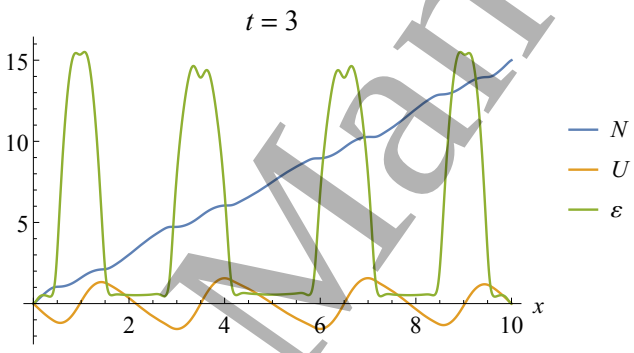


FIG. 21: Plot of solution to 1-D model at $t = 3$, illustrating final stage of evolution. The zonal flow field tilts in response to the intensity corrugation and merges into a quasisteady, persistent flow.

approximating the local spectrum $\varepsilon_{\mathbf{k}}$ by the local intensity ε, as the spectrum is likely to differ at early times.

E. Comparison to Ashourvan-Diamond model

Ashourvan and Diamond proposed[14,15] 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

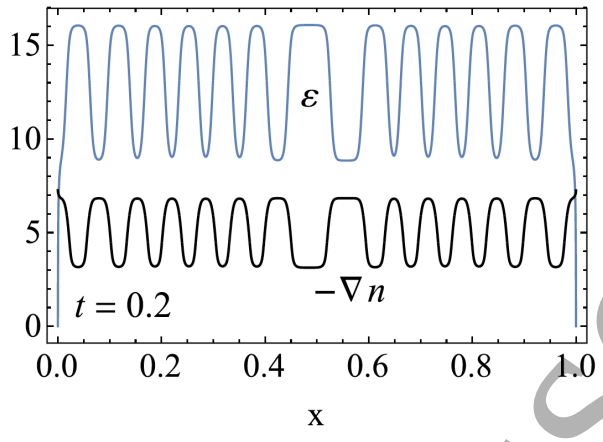


FIG. 22: Quasiperiodic features in the density and turbulent PE profiles in the AD model. This figure appeared previously in Ashourvan and Diamond (2017)[14].

Rhines mechanism[41] (in the spirit of Balmforth *et al.*[42]):

$$\ell_{mix}^2 = \frac{\ell_0^2}{\left(1 + \frac{\ell_0^2 (N' + V_y'')^2}{\varepsilon}\right)^{\kappa}}. \tag{61}$$

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, 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.[43] 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. 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.

V. 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[4,44]. 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.

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

$$\langle \partial_y \tilde{\phi} \nabla_\perp^2 \tilde{\phi} \rangle = \langle \partial_y \tilde{\phi} \partial_x^2 \tilde{\phi} \rangle + \langle \partial_y \tilde{\phi} \partial_y^2 \tilde{\phi} \rangle. \tag{A1}$$

The first term may be rewritten

$$\langle \partial_y \tilde{\phi} \partial_x^2 \tilde{\phi} \rangle = \left\langle \partial_x (\partial_x \tilde{\phi} \partial_y \tilde{\phi}) - \frac{1}{2} \partial_y \left((\partial_x \tilde{\phi})^2 \right) \right\rangle = \partial_x \langle \partial_x \tilde{\phi} \partial_y \tilde{\phi} \rangle, \tag{A2}$$

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

$$\langle \partial_y \tilde{\phi} \partial_y^2 \tilde{\phi} \rangle = \left\langle \frac{1}{2} \partial_y \left((\partial_y \tilde{\phi})^2 \right) \right\rangle = 0, \tag{A3}$$

again using periodicity. We are left with

$$\langle \partial_y \tilde{\phi} \nabla_\perp^2 \tilde{\phi} \rangle = \partial_x \langle \partial_x \tilde{\phi} \partial_y \tilde{\phi} \rangle. \tag{A4}$$

The RHS is the Reynolds force, as claimed.

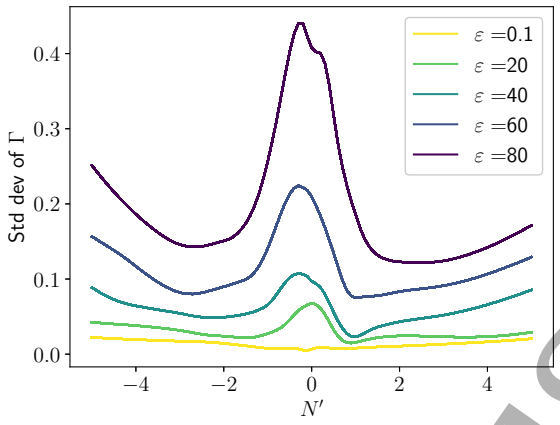


FIG. 23: Plot of the standard deviation among the ensemble of DNN models for the diffusive/diagonal part of the particle flux ($U = U' = U'' = 0$).

Appendix B: Error quantification

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. 23–25. 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.

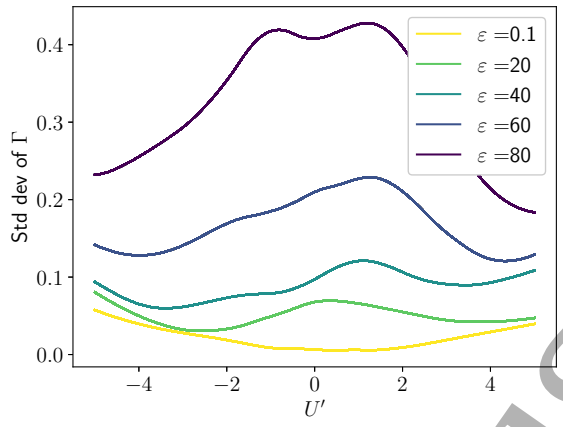


FIG. 24: Plot of the standard deviation among the ensemble of DNN models for the nondiffusive/off-diagonal part of the particle flux ($U = U'' = N' = 0$).

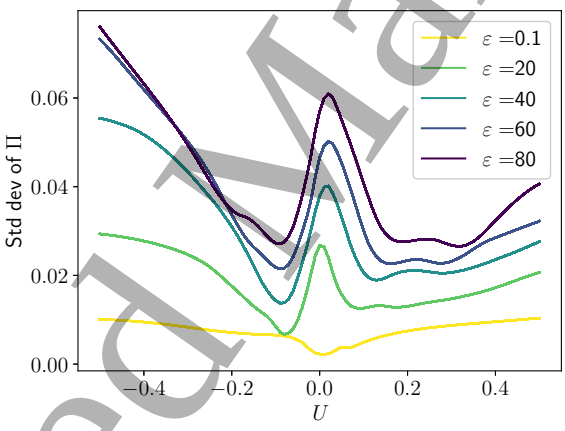


FIG. 25: Plot of the standard deviation among the ensemble of DNN models for the Reynolds stress, when $N' = 2$ and $U' = U'' = 0$.

ACKNOWLEDGMENTS

We acknowledge Arash Ashourvan, Norman Cao, Guilhem Dif-Pradalier, Ozgür Gürcan, and T. S. Hahm for useful discussions, many of which took place at the 2017 and 2019 Festivals de Théorie in Aix-en-Provence and the 2018 Chengdu Theory Festival. We also thank Olle Heinonen for suggestions on the numerical solution of the reduced 1-D model. This work used the Extreme Science and Engineering Discovery Environment (XSEDE)[45], which is supported by National Science Foundation grant number ACI-1548562, using the Comet cluster at the San Diego Su-

percomputing Center (SDSC) through allocation TG-PHY190014. It was supported by the U.S. Department of Energy, Office of Science, Office of Fusion Energy Sciences under Award Number DE-FG02-04ER54738.

REFERENCES

[1]W. Horton, Rev. Mod. Phys. **71**, 735 (1999).

[2]P. H. Diamond, S. Itoh, K. Itoh, and T. Hahm, Plasma Physics and Controlled Fusion **47**, R35 (2005).

[3]E.-j. Kim and P. H. Diamond, Physical Review Letters **90**, 185006 (2003).

[4]G. Dif-Pradalier, P. Diamond, V. Grandgirard, Y. Sarazin, J. Abiteboul, X. Garbet, P. Ghendrih, A. Strugarek, S. Ku, and C. Chang, Physical Review E **82**, 025401 (2010).

[5]G. Dif-Pradalier, G. Hornung, P. Ghendrih, Y. Sarazin, F. Clairet, L. Vermare, P. Diamond, J. Abiteboul, T. Cartier-Michaud, C. Ehrlacher, *et al.*, Physical Review Letters **114**, 085004 (2015).

[6]A. Ashourvan, R. Nazikian, E. Belli, J. Candy, D. Eldon, B. Grierson, W. Guttenfelder, S. Haskey, C. Lasnier, G. McKee, *et al.*, Physical Review Letters **123**, 115001 (2019).

[7]L. Prandtl, Z. Angew. Math. Meth. **5**, 136 (1925).

[8]B. B. Kadomtsev, *Plasma Turbulence* (Academic Press, 1965).

[9]A. Hasegawa and M. Wakatani, Physical Review Letters **50**, 682 (1983).

[10]M. Wakatani and A. Hasegawa, The Physics of Fluids **27**, 611 (1984).

[11]J. G. Charney, Journal of the Atmospheric Sciences **28**, 1087 (1971).

[12]A. Hasegawa and K. Mima, Phys. Rev. Lett. **39**, 205 (1977).

[13]G. I. Taylor and W. N. Shaw, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character **215**, 1 (1915).

[14]A. Ashourvan and P. H. Diamond, Physics of Plasmas **24**, 012305 (2017).

[15]A. Ashourvan and P. Diamond, Physical Review E **94**, 051202 (2016).

[16]Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).

[17]K. Hornik, Neural Networks **4**, 251 (1991).

[18]M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, Neural Networks **6**, 861 (1993).

[19]Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, in *Advances in neural information processing*

systems (2017) pp. 6231–6239.

[20]D. Rolnick, A. Veit, S. Belongie, and N. Shavit, arXiv preprint arXiv:1705.10694 (2017).

[21]R. A. Heinonen and P. H. Diamond, Phys. Rev. E Rap. Comm. (2020), (in press).

[22]R. Numata, R. Ball, and R. L. Dewar, Physics of Plasmas **14**, 102312 (2007).

[23]W. Dorland and G. W. Hammett, Physics of Fluids B: Plasma Physics **5**, 812 (1993).

[24]A. M. Dimits, G. Bateman, M. Beer, B. Cohen, W. Dorland, G. Hammett, C. Kim, J. Kinsey, M. Kotschenreuther, A. Kritz, *et al.*, Physics of Plasmas **7**, 969 (2000).

[25]B. Dudson, M. Umansky, X. Xu, P. Snyder, and H. Wilson, Computer Physics Communications **180**, 1467 (2009).

[26]G. E. Karniadakis, M. Israeli, and S. A. Orszag, Journal of computational physics **97**, 414 (1991).

[27]P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, Physics Reports **810**, 1 (2019).

[28]D.-A. Clevert, T. Unterthiner, and S. Hochreiter, arXiv preprint arXiv:1511.07289 (2015).

[29]S. Ioffe and C. Szegedy, arXiv preprint arXiv:1502.03167 (2015).

[30]D. P. Kingma and J. Ba, arXiv preprint arXiv:1412.6980 (2014).

[31]P. H. Diamond, S.-I. Itoh, and K. Itoh, *Modern Plasma Physics, Volume 1: Physical Kinetics of Turbulent Plasmas* (Cambridge University Press, 2010).

[32]P. W. Terry, Rev. Mod. Phys. **72**, 109 (2000).

[33]J. W. Cahn and J. E. Hilliard, The Journal of Chemical Physics **28**, 258 (1958), https://doi.org/10.1063/1.1744102.

[34]J. B. Parker, *Zonal flows and turbulence in fluids and plasmas*, Ph.D. thesis, Princeton University (2015).

[35]N. A. Bakas and P. J. Ioannou, Journal of Fluid Mechanics **682**, 332 (2011).

[36]K. Srinivasan and W. Young, Journal of the atmospheric sciences **69**, 1633 (2012).

[37]A. Smolyakov and P. Diamond, Physics of Plasmas **6**, 4410 (1999).

[38]D. Ruiz, J. Parker, E. Shi, and I. Dodin, Physics of Plasmas **23**, 122304 (2016).

[39]P. Goldreich and D. Lynden-Bell, Monthly Notices of the Royal Astronomical Society **130**, 125 (1965), https://academic.oup.com/mnras/article-pdf/130/2/125/8072031/mnras130-0125.pdf.

[40]L. O. Jay, in *Encyclopedia of Applied and Computational Mathematics*, edited by B. Engquist (Springer-Verlag, Berlin, 2015) pp. 817–826.

[41]P. B. Rhines, Journal of Fluid Mechanics **69**, 417 (1975).

[42]N. Balmforth, S. G. L. Smith, and W. Young, Journal of Fluid Mechanics **355**, 329 (1998).

[43]W. Guo, P. H. Diamond, D. W. Hughes, L. Wang, and A. Ashourvan, Plasma Physics and Controlled Fusion **61**, 105002 (2019).

[44]K. Ida, Z. Shi, H. Sun, S. Inagaki, K. Kamiya, J. Rice, N. Tamura, P. Diamond, G. Dif-Pradalier, X. Zou, *et al.*, Nuclear Fusion **55**, 013022 (2015).

[45]J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, *et al.*, Computing in Science & Engineering **16**, 62 (2014).