Characterizing the Reynolds number dependence of the chaotic attractor in two-dimensional turbulence with dimension-minimizing autoencoders

Andrew Cleary [*,†] and Jacob Page [‡]

School of Mathematics & Maxwell Institute for Mathematical Sciences, University of Edinburgh, Edinburgh EH9 3FD, United Kingdom

(Received 10 June 2025; revised 18 September 2025; accepted 21 October 2025; published 12 November 2025)

Deep autoencoder neural networks can generate highly accurate, low-order representations of turbulence. We design a family of autoencoders which are a combination of a "dense-block" encoder-decoder structure [Page *et al.*, J. Fluid Mech. **991**, A10 (2024)], an "implicit rank minimization" series of linear layers acting on the embeddings [K. Zeng *et al.*, Mach. Learn.: Sci. Technol. **5**, 025053 (2024)], and a full discrete + continuous symmetry reduction. These models are applied to two-dimensional turbulence in Kolmogorov flow for a range of Reynolds numbers $25 \leqslant \mathrm{Re} \leqslant 400$ and used to estimate the dimension of the chaotic attractor, $d_{\mathcal{A}}(\mathrm{Re})$. We find that the dimension scales like $\sim \mathrm{Re}^{1/3}$—much weaker than known bounds on the *global* attractor, which grow like $\mathrm{Re}^{4/3}$. In addition, two-dimensional maps of the latent space in our models reveal a rich structure not seen in previous studies, including multiple classes of high-dissipation events at lower Re, which guide bursting trajectories. We visualize the embeddings of large numbers of "turbulent" unstable periodic orbits, which the model indicates are distinct (in terms of features) from any flow snapshot in a large turbulent dataset, suggesting their dynamical irrelevance. This is in sharp contrast to their appearance in more traditional low-dimensional projections, in which they appear to lie within the turbulent attractor.

DOI: 10.1103/cdz2-858n

I. INTRODUCTION

Nonlinear partial differential equations (PDEs), such as the Navier–Stokes equations that govern turbulent fluid motion, are fundamental in modeling complex dynamics and often exhibit spatiotemporal chaos in parameter regimes of practical interest. Although the state spaces of PDEs are formally infinite dimensional, it is argued that solutions of the Navier–Stokes equations collapse onto a finite-dimensional inertial manifold, denoted here by $\mathcal{M}$, due to the dissipative action of viscosity [1–4]. The inertial manifold is a smooth, finite-dimensional manifold, within which the *global attractor* (which may be a complicated fractal set) is embedded [3,4]. The global attractor $\mathcal{A}_G$ is defined as the set of points in phase space that can be reached in arbitrarily long time from all initial conditions [5]. Two familiar properties of $\mathcal{A}_G$ are (i) it attracts the time forward map of all initial conditions $\boldsymbol{u}_0$, such that $\mathrm{dist}(\boldsymbol{f}^t(\boldsymbol{u}_0), \mathcal{A}_G) \to 0$ as $t \to \infty$, and (ii) it is an invariant set. Although the existence of inertial manifolds for the Navier–Stokes equations has not been proven [6], approximate inertial manifolds, which attract all orbits within a thin neighborhood of the global attractor, have been constructed in two spatial dimensions [7]. The inertial manifold encapsulates the essential features of the system, with an intrinsic dimension $d_{\mathcal{M}}$, typically much smaller than the dimension d_ω of the state space of the discretized PDE.

While the existence of an inertial manifold is not proven, the global attractor $\mathcal{A}_G \subset \mathcal{M}$ exists for the forced two-dimensional (2D) Navier–Stokes equations [2], and its dimension has been bounded from above [8,9] and below [10] (the latter in monochromatic Kolmogorov flow). The bounds are asymptotically tight (up to a logarithmic factor): The upper bound scales with a nondimensional Grashof number $\mathrm{Gr} := (\int_A |\mathbf{f}|^2 dA)^{1/2} L^2/\nu^2$ as $\mathrm{Gr}^{2/3}(1 + \ln \mathrm{Gr})^{1/3}$ [8,9], which depends on the forcing profile $\mathbf{f}$, box length L, and viscosity ν, while the lower bound scales like $\mathrm{Gr}^{2/3}$ [10] (This computation is based on the dimension of the unstable manifold of the laminar fixed point). Alternatively, a heuristic argument for the dimension of the inertial manifold has been made by determining the total number of excited modes in the inertial range of the (direct) turbulent cascade. This estimate scales like $N \sim (L/l_d)^2 \propto \mathrm{Gr}^{2/3}$ in two dimensions, where L is the largest length scale in the domain and l_d is the dissipative (Kraichnan) length scale [11]. This heuristic estimate was extended to $\mathrm{Gr}^{2/3}(\ln \mathrm{Gr})^{1/3}$ by incorporating a logarithmic correction to the energy spectrum to ensure a constant enstrophy flux in the direct cascade [12] and was shown to be consistent with the rigorous upper bound [13].

Our focus in this work is on generating low-dimensional embeddings for the chaotic attractor $\mathcal{A} \subset \mathcal{A}_G$ in two-dimensional Kolmogorov flow. $\mathcal{A}$ is defined as an invariant set which attracts an open set of initial conditions and is minimal—no proper subset of $\mathcal{A}$ exists, which satisfies the

[*]Contact author: andrew.cleary@ed.ac.uk

[†]Present address: Department of Mechanical Engineering, Johns Hopkins University, Baltimore MD 21218, USA.

[‡]Contact author: jacob.page@ed.ac.uk

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

first two conditions. The chaotic attractor presumably requires far fewer degrees of freedom than $\mathcal{A}_G$, which must contain all exact coherent structures and their unstable manifolds. To estimate the scaling $d_{\mathcal{A}} := \dim(\mathcal{A})$ with Reynolds number Re, we construct highly accurate autoencoder neural networks in an effort to find robust "latent" coordinate systems. Our objectives are twofold: (1) to estimate the Re dependence of $d_{\mathcal{A}}$ and (2) to use the latent coordinate system to map typical dynamical pathways through different vortical "events" and to understand the coverage (or lack of coverage) of these dynamics by large libraries of recently computed unstable periodic orbits [14,15].

Numerical computations for the chaotic attractor dimension (by computing Lyapunov exponents and invoking the Kaplan–Yorke conjecture [16]) give some indications as to a scaling with Re. The Lyapunov dimension of the attractor of Kolmogorov flow was reported to depend strongly on the wave number n chosen for the monochromatic forcing profile, and only weakly on $\mathrm{Re} \propto \mathrm{Gr}^{1/2}$ [17]. A follow-on study of Kolmogorov flow found that the Lyapunov dimension is bounded above by the number of degrees of freedom contained in scales larger than the forcing scale, $N_{LS} \propto n^2$ [18], though this study considered relatively weak turbulence with $16.4 \leqslant \mathrm{Re} \leqslant 46.2$ in our units. A scaling of the Lyapunov dimension $\sim \mathrm{Re}^{0.78}(n/k_0)^2$ was also computed numerically for forced two-dimensional turbulence with large-scale damping, where $k_0 = 2\pi/L$ [19], also demonstrating a linear dependence on N_{LS}.

Related to these direct numerical estimates, there has been a recent focus on using data-driven approaches to estimate the dimensionality of the inertial manifold of simpler chaotic systems. Linear approaches, such as principal component analysis (PCA) and its extensions (e.g., sparse PCA [20], Bayesian PCA [21]), tend to overestimate the dimension as they are limited to linear subspaces. Nonlinear techniques (e.g., deep autoencoders) address this limitation to more faithfully learn the structure of the inertial manifold. By examining how the model reconstruction error of the data changes with the dimensionality of the embedding space, autoencoder networks have been used to identify the inertial manifold dimension of the Kuramoto–Sivashinsky (KS) equation [22–24]. However, this approach requires a manual sweep over multiple latent dimensions, and there is no apparent drop-off in reconstruction error for more complex systems and higher spatial dimensions [25,26]. A recent study has proposed a more systematic approach to determine the dimensionality of the inertial manifold of the KS equation, using the so-called Implicit Rank Minimizing Autoencoder with Weight Decay (IRMAE-WD) [27]. The design of this architecture exploits the phenomenon that gradient-based optimization of deep *linear* neural networks leads to low-rank solutions (implicit rank minimization) [28], while the weight decay refers to an L_2 regularization on the weights of the network to further drive down the rank of the learned embedding. The same network was applied to both two-dimensional Kolmogorov flow with a forcing wave number of $n = 2$ at low Re [29] and two-dimensional Rayleigh–Bénard flow [30] to compute estimates for the dimensionality of these flows.

Many canonical fluid flows exhibit continuous and/or discrete symmetries in the governing equations which must be accounted for in the training of a nonlinear neural network. The simplest way to encourage the model to learn the symmetries of the equations is via data augmentation, where each symmetry is randomly applied to the training data after each epoch [14,31] to ensure a robust sampling of the symmetry groups. However, this approach does not guarantee equivariance of the model with respect to the symmetries. In this work, to explicitly reduce a continuous or discrete symmetry, each snapshot is mapped a symmetry-reduced subspace of the full state space. This preprocessing of the data restricts the model to learning a representation within this subspace [29,32]. The first Fourier mode method of slices can be used to explicitly reduce continuous symmetries of dynamical systems [33]. For discrete symmetries, we adapt a reduction technique that defines an indicator function based on Fourier modes to separate the state space of the governing equation into symmetric charts and maps all solutions into one of these charts, defining a "fundamental chart" (this was applied to the KS system in [32]). This method of "symmetry charting" has been applied to Kolmogorov flow with forcing wave number $n = 2$ [29], where an indicator function was defined on specific Fourier modes to define eight symmetry charts representing the eightfold discrete symmetries. Other work has explicitly accommodated various discrete symmetries via the use of many identical ("Siamese") networks to identify a fundamental domain [34].

Other studies have also considered the problem of simulating dynamics in the latent space of an autoencoder, motivated by the extreme computational requirements of direct numerical simulation of turbulence at high Re. Kolmogorov flow has been a popular testing ground for these efforts to date. For example, convolutional autoencoders have yielded a latent-space representation which revealed a clear separation between the high-dissipation bursting dynamics and the low-dissipation quiescent dynamics at moderate Reynolds number Re [14,35]. Autoencoders have also been combined with echo state networks as a reduced-order model to accurately predict the flow at different Re [36]. An imperfect reduced-order model for Kolmogorov flow has been complemented with a data-driven correction term to accurately predict extreme high-dissipation events [37].

In this work, we will combine convolutional "dense blocks" [14], the IRMAE-WD network approach [28,32] and a full symmetry reduction step via symmetry charting to estimate the scaling with Re of an upper bound to the chaotic attractor dimension of the $n = 4$ Kolmogorov flow. The combination of the IRMAE model with a full symmetry reduction yields a very rich and interpretable learned latent space that reveals physical insight to the dynamics of the flow. In Sec. II we formulate the equations of motion of two-dimensional Kolmogorov flow and describe the data generation and symmetry reduction. In Sec. III we describe the autoencoder architecture. In Sec. IV we report the estimated scaling of the chaotic attractor dimension with Re. In Sec. V we investigate the structure and modal decomposition of the learned embedding spaces of the models.

II. FLOW CONFIGURATION

A. Numerical simulations

We consider two-dimensional turbulence on a doubly periodic domain driven by a monochromatic body force in the streamwise direction (Kolmogorov flow). The out-of-plane vorticity $\omega = \partial_x v - \partial_y u$, where the velocity $\boldsymbol{u} = (u, v)$, evolves according to

$$\partial_t \omega + \boldsymbol{u} \cdot \nabla \omega = \frac{1}{\mathrm{Re}} \nabla^2 \omega - n \cos ny. \tag{1}$$

In this nondimensionalization, a length scale is chosen as the inverse of the fundamental wave number of the box $1/k^* = L^*/2\pi$, where $L^* = L_x^* = L_y^*$ (i.e., equal aspect ratio, the asterisk denoting dimensional units here). A timescale is chosen as $1/\sqrt{k^* \chi^*}$, where χ^* is the amplitude of the forcing in the momentum equation. These length and timescales define the Reynolds number for the configuration $\mathrm{Re} := \sqrt{\chi^*/k^{*3}}/\nu$, where ν is the kinematic viscosity. Throughout this work the forcing wave number is set to $n = 4$, as has been common in previous studies [15,35,38]. Equation (1) with periodic boundary conditions and $n = 4$ forcing waves is equivariant under the action of three symmetries:

(1) Continuous shifts in the streamwise (x) direction, $\mathscr{T}_s$: $\omega(x, y) \to \omega(x + s, y)$,

(2) Discrete shift-reflects by a half-wavelength in y, $\mathscr{S}$: $\omega(x, y) \to -\omega(-x, y + \pi/4)$, and

(3) Discrete rotations by π, $\mathscr{R} : \omega(x, y) \to \omega(-x, -y)$,

which we account for explicitly when estimating the dimension of the turbulent attractor.

Key measures of the flow used throughout this study are the total dissipation rate,

$$D := \frac{1}{\mathrm{Re}} \langle |\nabla \boldsymbol{u}|^2 \rangle_V, \tag{2}$$

and the total production rate,

$$I := \langle u \sin(ny) \rangle_V, \tag{3}$$

where the average over the volume $V = [0, 2\pi]^2$ is defined as

$$\langle \bullet \rangle_V := \frac{1}{(2\pi)^2} \iint_V \bullet \, d^2\boldsymbol{x}. \tag{4}$$

We consider this flow at a number of $\mathrm{Re} \in \{25, 30, 35, 40, 60, 80, 100, 200, 400\}$. Kolmogorov flow approaches an asymptotic regime [38,39] with $D \sim \mathrm{Re}^{-1/2}$ beyond $\mathrm{Re} \approx 50$. A variety of simpler states have been reported at lower Re [38]. The majority of Re values considered were sampled from the asymptotic chaotic regime. We will devote particular attention to the well-studied values $\mathrm{Re} = 40$ and $\mathrm{Re} = 100$.

We solve Eq. (1) using the spectral version of the `JAX-CFD` solver [40,41]. The velocity field is computed at each time step via solution of the Poisson equation $\nabla^2 \psi = -\omega$, where the stream function ψ is related to the induced velocity components via $u = \partial_y \psi$, $v = -\partial_x \psi$.

B. Datasets

For $\mathrm{Re} \leqslant 200$, training data was generated from 400 trajectories, each of length 10^3 advective time units, with an initial transient period of 50 advective time units discarded.

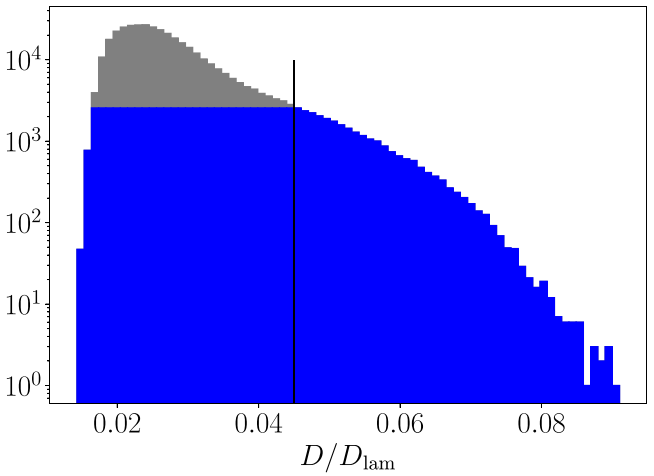


FIG. 1. The distributions of dissipation rate D at $\mathrm{Re} = 100$, normalized by the laminar dissipation rate $D_{\mathrm{lam}} = \mathrm{Re}/(2n^2)$, are shown for the full training dataset (gray) and the resampled training dataset (blue).

At $\mathrm{Re} = 400$ we instead computed 4000 trajectories each of length 10^2 (to enable parallelization over trajectories and avoid excessive computation time at high resolution). Data was sampled every advective time unit ($\Delta t^* = 1/\sqrt{k^* \chi^*}$ in dimensional units), resulting in a dataset of 4×10^5 individual snapshots for each Re value considered. A spatial resolution of 128^2 was sufficient for $\mathrm{Re} < 200$, while trajectories at $\mathrm{Re} = 200$ and 400 were computed with 256^2 and 512^2 grid-points, respectively, before being downsampled to 128^2 for input to the autoencoders. We reserved 10% of the training data as a "validation" dataset, which is not used to update the model weights but instead is used to track convergence metrics throughout training. Test datasets of 5×10^4 snapshots for each Re were generated in the same manner.

1. Resampling

Training models on data sampled from the turbulent attractor, as described above, lead to low reconstruction errors for low-dissipation-rate snapshots but also high errors for high-dissipation-rate snapshots as observed in previous studies [14,35]. This is because (1) low-dissipation snapshots are less spatially complex and (2) high-dissipation events are rare events in Kolmogorov flow [38,42]—the dataset is dominated by low-dissipation events (see Fig. 1). One approach that has previously been used to alleviate this effect is the introduction of a loss term involving the square of the vorticity field [14] to add additional penalization to higher-dissipation events. Here, we instead introduce a resampling strategy to alter the distribution of training data over dissipation and artificially increase the frequency of high-dissipation events. To do this, the distribution of dissipation rate is first partitioned into "low" and "high"-dissipation events, defined by a threshold value D_c. For example, at $\mathrm{Re} = 100$, this value is set to $D_c = 0.045 D_{\mathrm{lam}}$ (the laminar dissipation rate is $D_{\mathrm{lam}} = \mathrm{Re}/2n^2$). All data above the threshold value is sampled from the original training dataset. The dataset is then binned according to dissipation rate, with bin widths of $\Delta D = 0.001 D_{\mathrm{lam}}$ for $\mathrm{Re} < 200$, $\Delta D = 0.0005 D_{\mathrm{lam}}$ for $\mathrm{Re} = 200$, and $\Delta D = 0.0001 D_{\mathrm{lam}}$ for

Re = 400. To smoothly reduce the number of low-dissipation events in the dataset, a maximum of n_l snapshots are then randomly sampled from each bin below the threshold D_c, where n_l is set by the number of samples in the first high-dissipation bin $[D_c, D_c + \Delta D]$. The modified distribution created by this resampling strategy is shown in Fig. 1 for Re = 100.

The parameter D_c must balance two key considerations: that (1) the network is exposed to a relatively high number of high-dissipation snapshots and that (2) a sufficiently large training dataset is retained after resampling from the 4×10^5 original snapshots at each Re. These are competing conditions, as the size of the resampled dataset can be increased by lowering D_c but with a reduced weighting given to the high-dissipation snapshots. The thresholds, in order of increasing Re, are set at $D_c/D_{\text{lam}} \in \{0.16, 0.1415, 0.1255, 0.15, 0.125, 0.0725, 0.045, 0.0125, 0.0045\}$. As a result of the resampling, the size of the training dataset is reduced, with a total number of snapshots $N_S \approx (20, 40, 50, 60, 75, 90, 100, 125, 130) \times 10^3$ (ordered in increasing Re).

2. Symmetry reduction

As described in Sec. II A, Kolmogorov flow is equivariant under continuous shifts in the streamwise direction, under discrete shift-reflects by a half-wavelength in y, and under rotations by π. We account for all symmetries explicitly in a preprocessing step before training the neural networks. The continuous translational symmetry is reduced by the first Fourier mode method of slicing [33]. In this method a phase ϕ is computed such that the $(k_x, k_y) = (1, 0)$ Fourier mode of the translated state $\mathcal{T}_{-\phi}\omega$ is purely real.

For the discrete symmetries, we adapt the symmetry charting approach introduced in [29] for $n = 2$ Kolmogorov flow. This approach partitions the full state space into symmetrically related charts and assigns each chart a unique indicator $\mathscr{I}$. We select a single chart to which we map all states by combinations of $\mathscr{S}$ and $\mathscr{R}$, which is done by looking at specific terms in the Fourier transform of the vorticity field:

$$\omega(x, y) = \sum_{k_x} \sum_{k_y} \hat{\omega}(k_x, k_y) e^{i(k_x x + k_y y)}.$$

The actions of the three symmetry operations on the Fourier modes are

$$\mathcal{T}_s : \hat{\omega}(k_x, k_y) \to \hat{\omega}(k_x, k_y) e^{-ik_x s}, \tag{5}$$

$$\mathscr{S} : \hat{\omega}(k_x, k_y) \to -\hat{\omega}(-k_x, k_y) e^{ik_y \pi / n}, \tag{6}$$

$$\mathscr{R} : \hat{\omega}(k_x, k_y) \to \overline{\hat{\omega}}(k_x, k_y), \tag{7}$$

where $\overline{\bullet}$ denotes complex conjugation. It is clear from (5) that the action of $\mathcal{T}_s$ only changes modes with $k_x \neq 0$. However, $\mathscr{R}$ changes all modes, while $\mathscr{S}$ changes all modes except for modes $[0, (2l + 1)n]$, where $l \in \mathbb{Z}$. As $\mathcal{T}_s$ does not affect modes with $k_x = 0$, we will use specific $k_x = 0$ modes to specify the indicators of the discrete symmetries and reduce the discrete symmetries before the translational symmetry. Otherwise, if the discrete symmetries are reduced after the translational symmetry, the (1,0) mode is no longer guaranteed to be purely real.

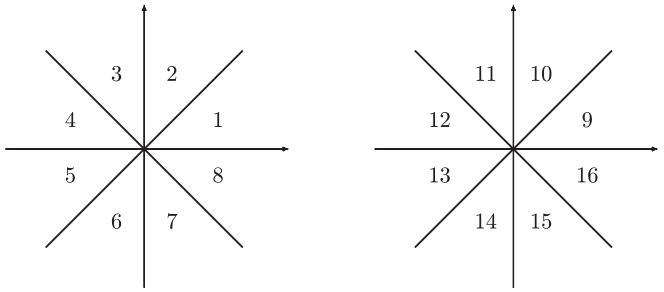


FIG. 2. Sixteen symmetry charts of $n = 4$ Kolmogorov flow, for each of the 16 discrete symmetric copies of a state. The x and y axes denote $\hat{\omega}_R(0, 1)$ and $\hat{\omega}_I(0, 1)$, respectively. (Left) $\hat{\omega}_I(0, 4) > 0$ and (right) $\hat{\omega}_I(0, 4) < 0$.

We consider the (0, 1) mode to diagnose the shift-reflect symmetry, splitting the complex plane spanned by the real $\hat{\omega}_R(0, 1)$ and imaginary $\hat{\omega}_I(0, 1)$ components of this mode into octants, representing the eight shift-reflected copies generated by (6). We consider the (0, 4) mode for the rotational symmetry, splitting this complex plane into upper and lower halves, i.e., defined by the sign of $\hat{\omega}_I(0, 4)$. Together, this decomposition results in 16 indicators as indicated in Fig. 2, representing the 16 symmetric copies generated by both (6) and (7). To reduce the discrete symmetries of a state, we select the "fundamental" domain as $\mathscr{I} = 1$, which is defined by $0 < \arg \hat{\omega}(0, 1) < \pi/4$ and $\hat{\omega}_I(0, 4) > 0$.

The required combination of symmetry operations to map a state from any $\mathscr{I}$ to $\mathscr{I} = 1$ is tabulated in Table I. In practice, the symmetry reduction of a snapshot proceeds as follows: (i) identify the domain $\mathscr{I}$ of the snapshot from the (0, 1) and (0, 4) modes as in Fig. 2, (ii) apply the appropriate symmetry operations from Table I to the snapshot, and (iii) reduce the translational symmetry with the first Fourier mode method of slicing. Generalization of this approach to any number of forcing waves n is relatively straightforward—it requires the partitioning of the complex plane spanned by $\hat{\omega}_R(0, 1)$ and

TABLE I. Symmetry operations to map all 16 symmetry charts of $n = 4$ Kolmogorov flow to the fundamental chart $\mathscr{I} = 1$.

$\mathscr{I}$	Action
1	$\mathbb{I}$
2	$\mathscr{S}^3$
3	$\mathscr{S}^6$
4	$\mathscr{S}$
5	$\mathscr{S}^4$
6	$\mathscr{S}^7$
7	$\mathscr{S}^2$
8	$\mathscr{S}^5$
9	$\mathscr{S}^5\mathscr{R}$
10	$\mathscr{S}^2\mathscr{R}$
11	$\mathscr{S}^7\mathscr{R}$
12	$\mathscr{S}^4\mathscr{R}$
13	$\mathscr{S}\mathscr{R}$
14	$\mathscr{S}^6\mathscr{R}$
15	$\mathscr{S}^3\mathscr{R}$
16	$\mathscr{R}$

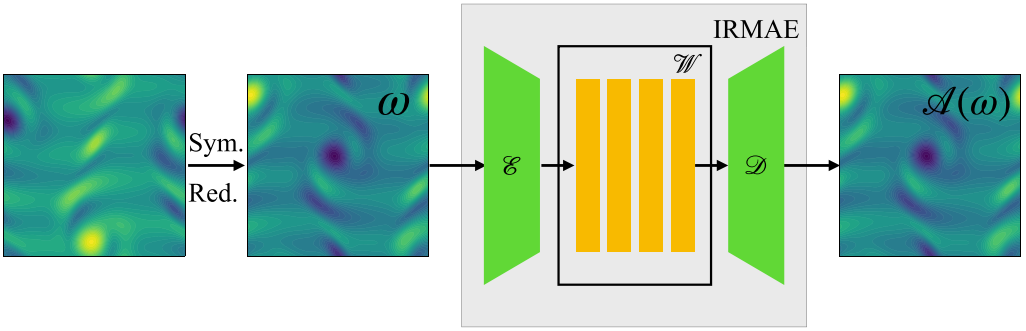


FIG. 3. Schematic of the data and neural network pipeline used in this study. Symmetries are first reduced for each snapshot ω, before being passed to the IRMAE network. The implicit rank minimizing layers (orange) are linear fully connected layers, while the encoder $\mathcal{E}$ and decoder $\mathcal{D}$ are fully described in Appendix A.

$\hat{\omega}_I(0, 1)$ into n equally sized sectors, along with consideration of the sign of $\hat{\omega}_I(0, n)$ to diagnose rotation.

III. NETWORK ARCHITECTURE

We construct low-dimensional representations of the turbulent attractor in Kolmogorov flow by combining the DenseNet convolutional autoencoder architecture from a previous study [14] (hereafter PHBK) with the implicit rank minimization and weight decay introduced by the IRMAE-WD architecture [27,29]. Each deep convolutional autoencoder $\mathscr{A}$ seeks to reconstruct its input such that

$$\mathscr{A}(\omega) \equiv [\mathscr{D} \circ \mathscr{W} \circ \mathscr{E}](\omega) \approx \omega, \tag{8}$$

where the encoder $\mathscr{E} : \mathbb{R}^{N_x \times N_y} \to \mathbb{R}^{d_z}$ maps the input vorticity snapshot to a low-dimensional representation, $\mathscr{W} : \mathbb{R}^{d_z} \to \mathbb{R}^{d_z}$ represents a series of n fully connected, equally sized linear layers (pure matrix multiplication) within the embedding space such that $\mathscr{W} = \mathbf{W}_1\mathbf{W}_2 \cdots \mathbf{W}_n$ and $\mathbf{W}_i \in \mathbb{R}^{d_z \times d_z}$, and the decoder $\mathscr{D} : \mathbb{R}^{d_z} \to \mathbb{R}^{N_x \times N_y}$ maps the embedding back to a vorticity snapshot. A high-level schematic of this architecture and the symmetry reduction pre-processing step is presented in Fig. 3.

The structures of the encoder $\mathscr{E}$ and decoder $\mathscr{D}$ considered here are adapted from PHBK, with so-called "dense blocks" [43,44] instead of single convolutional layers. The dense blocks allow for much richer feature maps at each scale throughout the autoencoder. Full details on the architecture used here are given in Appendix A; see Ref. [45] for the full source code. The key addition to the architecture is the insertion of fully connected linear layers into the bottleneck of the autoencoder. The implicit rank minimization of these linear layers drives the output of $\mathscr{W}(\mathscr{E}(\omega))$ to be of minimal rank [28]. The densely connected linear layers to the autoencoder bottleneck do not increase the expressivity of the network, as these layers can be expressed as a multiplication by a square matrix. However, the number of parameters increases to $\sim 6.21 \times 10^6$ trainable parameters due to the insertion of these linear layers, nearly triple the $\sim 2.15 \times 10^6$ trainable parameters in the comparable models considered in PHBK.

Although more complex (generative) approaches have been shown to learn informative latent representations of chaotic dynamics [46,47], the IRMAE network, while fully deterministic, has been shown to result in latent (or embedding) spaces with similar favorable properties. For example, comparable performance to variational autoencoders has been demonstrated in tasks such as smoothly interpolating in the latent space and generating new samples from random noise [28]. Furthermore, while the latent spaces of variational autoencoders are typically full rank, the IRMAE networks benefit from approximately minimal-rank latent spaces, which are key in estimating the dimensionality of the chaotic attractor.

The networks are trained to minimize the standard mean-squared-error loss function,

$$\mathscr{L} = \frac{1}{N_S} \sum_{j=1}^{N_S} \|\mathscr{A}(\omega_j) - \omega_j\|^2, \tag{9}$$

where each ω_j is a snapshot from the training dataset, using an AdamW optimizer [48] with an initial learning rate of 10^{-4} and the default weight decay coefficient of 0.004. This optimizer decouples the weight decay from the adaptive gradient updates, as including an explicit L_2 regularization term on the trainable parameters in (9) is not effective with Adam-style optimizers [49] nor identical to weight decay [48]. The batch size is set to 8 for $\mathrm{Re} \leqslant 40$, 12 for $\mathrm{Re} = 60$, and 16 for $\mathrm{Re} \geqslant 80$. Setting relatively small batch size increases the total training time but also increases the total number of gradient-based optimization steps, which we found to be effective in driving down the rank of the learned embedding. The networks are trained with the AdamW optimizer for at least 1000 epochs for each Re. Inspired by previous studies implementing a two-optimizer training process [50], the best model to that point is then further tuned using the AdaGrad optimizer [51] for a further 1000 epochs. A brief hyperparameter analysis over the initial learning rate and batch size is presented in Appendix C. The model size and associated training time (2 weeks of GPU time per model reported here) prohibited us from performing an extensive hyperparameter sweep over all Re, and better performance may well be possible with alternative settings.

Alternative loss functions, for instance, asserting that the time-evolved reconstructed vorticity fields must match the time-evolved ground truth [52] and spectral regularization terms which penalize poor energy spectrum reconstruction

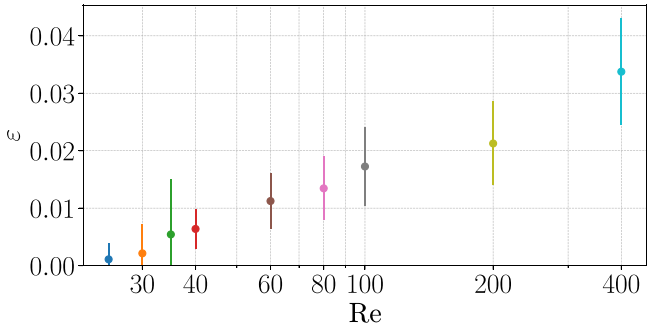


FIG. 4. The average reconstruction error ε of the model against Re. A bottleneck dimension of $d_z = 1024$ is chosen for each Re. The error bars at each Re are defined by one standard deviation of the reconstruction error.

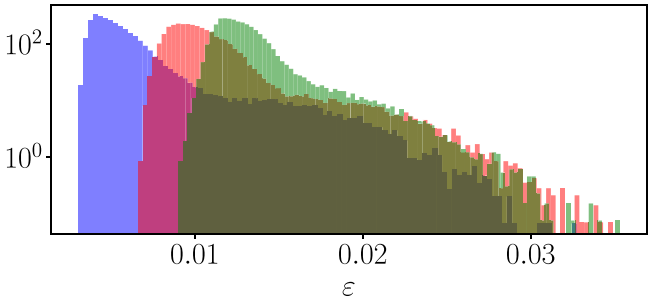


FIG. 5. The distribution of the reconstruction error ε at Re = 40 is shown for the best model trained with data augmentation and no symmetry reduction (green), the best model from the first AdamW optimizer stage on symmetry-reduced data (red), and the best model from the second Adagrad optimizer stage on symmetry-reduced data (blue).

[53], were also considered here but did not result in any substantial improvements to the reconstruction error. We also attempted to fit a second-stage autoencoder network, sharing the same latent representation of the data, to the residual of the reconstructed vorticity fields, motivated by the architecture used to perform machine precision reconstruction of a single turbulent snapshot [54]. However, the error did not appear to have any structure to learn and the second-stage autoencoder was unable to reduce the error any further.

IV. DIMENSION ESTIMATION

In this section we first report the reconstruction error of our model across all Re considered. We then present the estimated scaling for (an upper bound on) the dimension of the chaotic attractor with Re.

A. Reconstruction error

We report the relative error for each snapshot ω_j in the test datasets,

$$\varepsilon_j := \frac{\|\omega_j - \mathscr{A}(\omega_j)\|}{\|\omega_j\|}, \tag{10}$$

where the norm is defined as $\|\omega_j\| := \sqrt{(1/4\pi^2)\int\int \omega_j^2 d^2\mathbf{x}}$. The average test error for all Re examined is reported in Fig. 4, with error bars defined by one standard deviation. As expected, the error increases as Re is increased, from $\varepsilon \sim 0.5\%$ at Re = 40, up to $\varepsilon \sim 3.5\%$ at Re = 400. These models are more than twice as accurate as those described previously in Ref. [14].

The role of some of the features introduced in the data preprocessing/network training approach is assessed in Fig. 5. The distribution of errors from the best model from the first stage of training with the AdamW optimizer (red) is notably shifted compared with the best model from the second stage of training with the AdaGrad optimizer (blue). Heuristically, this two-optimizer training process first explores the training loss landscape via the inherent momentum and bias correction of AdamW, before AdaGrad's feature-adapted learning rate allows it to settle more deeply into the best local minimum encountered in the first training stage. To quantify the impact of the full symmetry reduction preprocessing, we train on a dataset containing all symmetries in which data augmentation is applied at each training epoch by a random application of the discrete and continuous symmetry operations. Even with the data augmentation, the reconstruction error of this model (green) is notably larger than the model trained on fully reduced data, as shown in Fig. 5. We shall also see in Sec. V that explicitly factoring out the symmetries leads to a rich latent-space structure.

The distribution of the error and correlation with dissipation rate is examined for a sample of the Re values considered in Fig. 6. For the Re = 40 results, we also compare the performance of identical architectures trained on dissipation-resampled (color) or not resampled (gray) data at Re = 40. For a fair comparison, a similar number of training snapshots was used in both cases. The resampling strategy considerably improves the error of the high-dissipation events and, surprisingly, also slightly improves the error of the low-dissipation events. The reconstruction error is still slightly correlated with dissipation rate, especially at higher Re.

We report the test-averaged enstrophy spectra $\Omega(k)$ generated by the model at Re = 40, 100, 200, and 400 in Fig. 7. We have replaced the traditional single-output layer in our architecture by a sequence of linear convolutional layers of decreasing kernel sizes from (16, 16) to (6, 6) (full details in Appendix A). This "output stream" of layers was motivated by the empirical observation that the reconstructed enstrophy spectrum with a single output convolutional layer plateaued incorrectly at high wave numbers. This plateau is evident in the top left panel of Fig. 7, where the red residual spectrum is generated by an identical fully trained architecture with a single linear output layer of filter size (8, 8) instead of the output stream. Empirically, the reduction in filter size through the stream tends to sequentially refine at finer spatial scales. Comparing the enstrophy spectra of the output with and without the output stream at Re = 40, we see a notable improvement at the largest spatial scale and a slight improvement at the very small spatial scales. We found that using any smaller filter sizes resulted in an oversmoothing of the reconstructed vorticity fields. This plateau at the high wave numbers is increasingly absent at higher Re and virtually

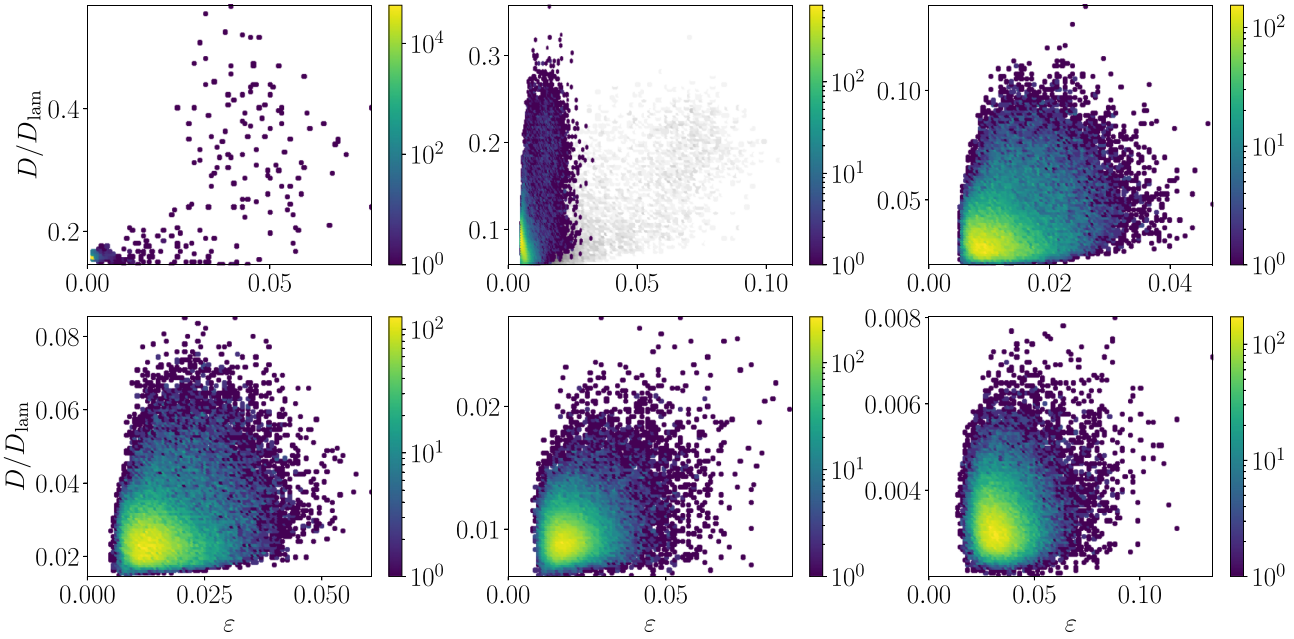


FIG. 6. Histograms of the test datasets at (top) Re = 25, 40, 80 and (bottom) Re = 100, 200, 400 from left to right, comparing the snapshot dissipation rate D normalized by the laminar dissipation rate $D_{\text{lam}} = \text{Re}/(2n^2)$ to the reconstruction error ε of the snapshot. The colored histograms in each panel correspond to models trained with resampled training data, while the gray histogram at Re = 40 (also on a logarithmic scale) corresponds to the model trained on the same number of nonresampled data.

nonexistent at Re = 400. Instead, the reconstructed fields are overly smoothed at the higher Re values of 200 and 400.

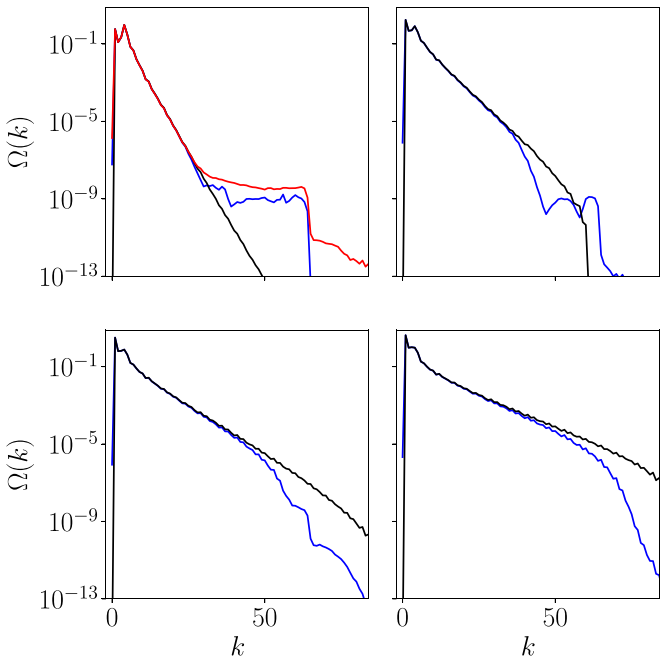


FIG. 7. The enstrophy spectrum $\Omega(k)$ of the (blue) reconstructed snapshots $\mathcal{A}(\omega)$ and (black) true snapshots ω at (top left) Re = 40, (top right) Re = 100, (bottom left) Re = 200, and (bottom right) Re = 400. The red line indicates the enstrophy spectrum at Re = 40 of a model trained with a single linear convolutional output layer of filter size (8, 8).

B. Attractor dimension

Here we present the main result of this study—the scaling with Re of the estimated dimension of the chaotic attractor. Implicit rank minimization drives the rank of the embedding space to an approximately minimal-rank solution, which we use as (an upper bound for) our numerically estimated dimension. To estimate this minimal rank, we perform a singular value decomposition (SVD) on the latent data matrix $\mathbf{Z}\mathbf{Z}^T$ [27], where $\mathbf{Z} = [\dots z_i \dots]$ and $z_i = \mathscr{W}(\mathscr{E}(\omega_i))$ for all ω_i in the test dataset. The SVD of $\mathbf{Z}\mathbf{Z}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T$ computes the singular vectors $\mathbf{u}_j \in \mathbb{R}^{d_z}$ where $\mathbf{U} = [\mathbf{u}_0 \dots \mathbf{u}_{d_{z-1}}]$ (the left singular vectors of $\mathbf{Z}$) and singular values $\sigma_j \in \mathbb{R}$ where $\mathbf{\Sigma} = \text{diag}(\sigma_0, \dots, \sigma_{d_{z-1}})$ (the square of the singular values of $\mathbf{Z}$). The number of non-negligible singular values in this spectrum is the estimate for $d_{\mathcal{A}}$ in the fully symmetry-reduced subspace. Each of these singular vectors represents some important structure of the original data in the latent space, which will be investigated in more detail in Sec. V.

The singular values for each Re considered are reported in the top panel of Fig. 8. In all cases there is a sharp drop before σ_j becomes negligibly small at a critical index $j = c$ (defined as the first index where $\sigma_c < 10^{-8}$), corresponding to the rank of the learned, fully-symmetry-reduced latent space. To compute the estimate for $d_{\mathcal{A}}$ in full state space, which we denote by $\tilde{d}_{\mathcal{A}}$, we add the continuous symmetry back into the estimate such that $\tilde{d}_{\mathcal{A}} = c + 1$ (reducing the discrete symmetries reduces the size of the state space, but not the dimension). We computed two estimates for $\tilde{d}_{\mathcal{A}}$ at each Re by training two randomly initialized autoencoders from scratch. These estimates were found to be robust to the training—they differed only slightly between models at all Re. The smallest estimate at each Re is reported in the bottom panel of Fig. 8.

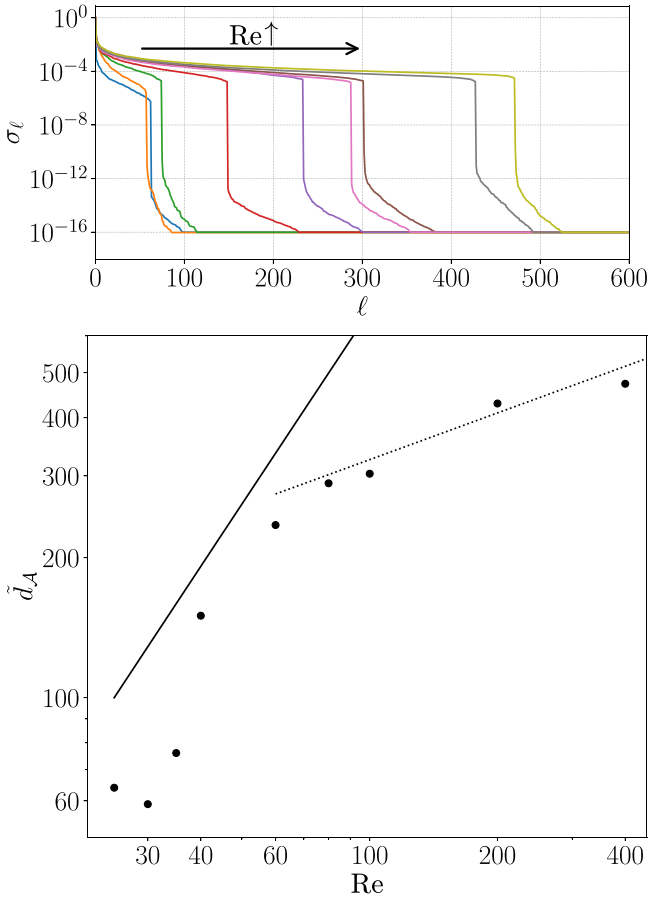


FIG. 8. (Top) Singular values σ_i in the singular value decompositions of the covariance of the latent data matrix at Re = 25 (—), 30 (—), 35 (—), 40 (—), 60 (—), 80 (—), 100 (—), 200 (—), and 400 (—). (Bottom) The estimated upper bound for the dimension of the turbulent attractor $\tilde{d}_\mathcal{A}$ against Re. The bold line shows the rigorous upper bound scaling $\sim \mathrm{Re}^{4/3}[1 + \ln(2^{5/2}\pi^3\mathrm{Re}^2)]^{1/3}$ and the dotted line shows the scaling $\sim \mathrm{Re}^{1/3}$.

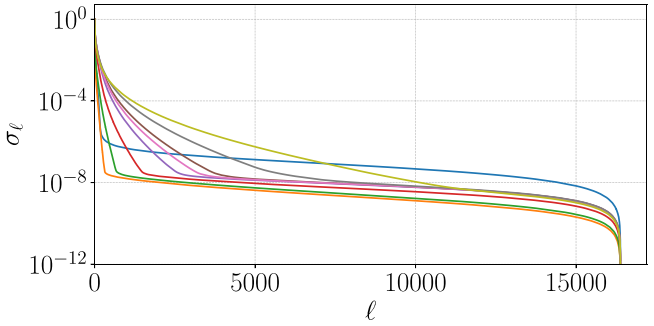


FIG. 9. Principal component analysis values σ_ℓ for the symmetry-reduced test datasets at Re = 25 (—), 30 (—), 35 (—), 40 (—), 60 (—), 80 (—), 100 (—), 200 (—), and 400 (—).

The nondimensional Grashof number Gr in the rigorous upper bound [8,9] of the *global* attractor dimension is related to the present definition of Re, with monochromatic $n = 4$ Kolmogorov forcing, by $\mathrm{Gr} = 2^{5/2}\pi^3\mathrm{Re}^2$. This upper bound then scales with Re like $b\mathrm{Re}^{4/3}[1 + \ln(2^{5/2}\pi^3\mathrm{Re}^2)]^{1/3}$ and is indicated by the bold line in the bottom panel of Fig. 8 (b is an unspecified constant). This bound and the lower bound in [10] for the global attractor both scale with the number of degrees of freedom in a two-dimensional domain ($\mathrm{Re}^{4/3}$ is proportional to the square of the domain size L over the Kraichnan scale) with a logarithmic correction.

In contrast, our data-driven estimate for the dimensionality of the *chaotic* attractor explored by long turbulent trajectories exhibits a much weaker scaling at high Re. While we observe some nonmonotonic behavior in the bottom panel of Fig. 8 at lower Re, there is a sharp transition beyond Re = 40 beyond which an approximate scaling $d_\mathcal{A} \sim \mathrm{Re}^{1/3}$ is observed in the range $\mathrm{Re} \in [60, 400]$, identified with the dotted line in the figure. Although this scaling was estimated over this limited range of Re values, earlier work [38,39] indicates that an "asymptotic" regime (a clear power law scaling of dissipation with Re) is established for $\mathrm{Re} \gtrsim 60$ and extends to at least Re = 1000, which would suggest that our dimension scaling may well also continue to be robust beyond Re = 400. The estimate $d_\mathcal{A} \sim \mathrm{Re}^{1/3}$ is a weaker scaling than a previous result of $\sim \mathrm{Re}^{0.78}$, which was computed via the Kaplan–Yorke conjecture [19]. However, this flow configuration was forced by a different profile and was damped at large scales, making a direct comparison challenging.

As a point of comparison for the nonlinear dimension estimation by the IRMAE networks, we also performed a linear dimensionality reduction approach on the symmetry-reduced test datasets using principal component analysis (PCA) [55]. The singular values from a PCA performed at each Re are reported in Fig. 9. There is no sharp cutoff in the singular values, especially at the higher Re, so it is not possible to meaningfully extract a dimension at each Re using this approach.

One caveat with our estimates is that our models do not perfectly reconstruct their inputs, and this error does increase with Re. However, the markedly weaker scaling with Re is consistent with the formation of coherent structures which are associated with a subset of wave-number pathways in a Fourier representation. We also examined the relative error with respect to a reference *trajectory* as the model predictions are unrolled in time and found that the errors remain roughly constant for longer than a Lyapunov time (we compared at Re = 100 where this time has been estimated at $T_\lambda = 3.6$ by [56]). Another caveat is we also assume that we have sufficiently sampled the chaotic attractor with our test datasets of 50 000 snapshots at each Re. We now explore the representation of the turbulence learned in these models by decoding individual components from the SVD.

V. LATENT-SPACE STRUCTURE

In this section we investigate the structure of the latent space learned by the model at various values of Re. By explicitly factoring out the symmetries of Kolmogorov flow, the model does not have to learn implicit representations of these symmetry groups. This results in a much richer representation of the dynamics in the fundamental symmetry chart that can be explicitly related to all symmetric copies.

A. Modal structures

We first examine the contribution to the vorticity field reconstruction from each of the singular vectors (modes) $\mathbf{u}_\ell$ in the SVD of $\mathbf{Z}\mathbf{Z}^T$. To do this, we project low- and high-dissipation snapshots onto the $\ell = 0$ mode and combinations of this mode with higher modes. The projection operator onto mode ℓ in the embedding space for some vorticity field ω is defined by

$$\mathcal{P}_\ell(\omega) = \langle \mathbf{u}_\ell, \mathscr{W}(\mathscr{E}(\omega)) \rangle, \tag{11}$$

where $\langle \cdot, \cdot \rangle : \mathbb{R}^{d_z} \times \mathbb{R}^{d_z} \to \mathbb{R}$ denotes the inner product in the embedding space. The projection onto the non-negligible modes $\ell < c$ defines the coordinates of ω in minimal-rank coordinates,

$$\phi(\omega) := (\mathcal{P}_0(\omega), \ldots, \mathcal{P}_{c-1}(\omega)) \in \mathbb{R}^c. \tag{12}$$

To visualize the physical structures spanned by each mode, we decode the projections onto each mode back to physical space. We add the projection onto the $\ell = 0$ mode when projecting onto $\ell > 0$ modes, i.e., $\mathcal{D}(\mathcal{P}_\ell(\omega)\mathbf{u}_\ell + \mathcal{P}_0(\omega)\mathbf{u}_0)$, as otherwise the output $\mathcal{D}(\mathcal{P}_\ell(\omega)\mathbf{u}_\ell)$ is not physically realistic (for a similar discussion, see Refs. [14,35]. Then we subtract $\mathcal{D}(\mathcal{P}_0(\omega)\mathbf{u}_0)$ from the decoded field in an effort to isolate and interpret the contribution from the $\ell > 0$ mode in physical space.

The contributions $\mathcal{D}(\mathcal{P}_0(\omega)\mathbf{u}_0)$ from mode 0 and $\mathcal{D}(\mathcal{P}_\ell(\omega)\mathbf{u}_\ell + \mathcal{P}_0(\omega)\mathbf{u}_0) - \mathcal{D}(\mathcal{P}_0(\omega)\mathbf{u}_0)$ from modes $\ell = 1, \ldots, 6$ are shown in Fig. 10 for a selection of snapshots spanning the full range of dissipation rates at $\mathrm{Re} = 40$ and $\mathrm{Re} = 100$. The same modes for $\mathrm{Re} = 200$ and $\mathrm{Re} = 400$ are reported in Fig. 11. At all Re considered, the $\ell = 0$ mode bears some resemblance to the time-averaged vorticity state in the symmetry-reduced subspace across the full range of dissipation rates. This structure is also visually similar to many low-dissipation *exact* solutions in this flow (traveling waves and relative periodic orbits, see [35,38,57] – some examples at $\mathrm{Re} = 40$ are also included in Appendix B). There is a weaker projection onto this state for the high-dissipation snapshots at $\mathrm{Re} = 40$, whereas the projection at the higher $\mathrm{Re} \in \{100, 200, 400\}$ is approximately equal across all snapshots presented. Similar to previous studies [14,35], the $\ell = 0$ mode is central to the latent representation in that the singular value σ_0 is two orders of magnitude larger than σ_1 at $\mathrm{Re} = 40$. The $\ell = 1$ mode also captures similar structures at both $\mathrm{Re} = 40$ and $\mathrm{Re} = 100$ and is dominated by paired diagonal elongated vortex cores of opposite circulation. This mode is visually similar to the most unstable eigenfunction of many low-dissipation exact solutions which resemble the mean flow state (see Fig. 18 in Appendix B).

At the higher $\mathrm{Re} = 200$ and $\mathrm{Re} = 400$, the representation changes somewhat; there is less evidence of the Kolmogorov forcing profile in the decodes of the dominant SVD vectors. Instead, many of the modes capture local features in the vorticity, including dipole/tripolar structures in different orientations (in particular, note modes beyond $\ell \geqslant 3$.

One striking aspect of the decodes at all Re is the sparsity in the representation; some modes are not active for the input snapshots. For example, consider modes $\ell = 1$ and $\ell = 2$ in the left column of Fig. 10. The projections onto these modes $\mathcal{P}_\ell$ is plotted on a logarithmic scale in the bottom row of Fig. 10, demonstrating that $\mathcal{P}_2$ is two orders of magnitude smaller than $\mathcal{P}_1$. This is suggestive of a strong separation of modes for different dynamical processes but is also likely due in part to the discrete symmetry reduction. The mapping to a fundamental chart is based on the first Fourier mode in the vertical. This can mean that snapshots on a trajectory can suddenly go through a "jump" to another chart that suddenly shifts the local vortical features when symmetry reduction is performed (the first Fourier mode may not be apparent by eye, particularly in high-dissipation snapshots), and the autoencoder therefore must learn to represent small scale vortices in different areas of the symmetry-reduced snapshots. This is perhaps apparent in the $\ell = 3$ mode at $\mathrm{Re} = 400$ (Fig. 11) with a dipole in the upper corners of the decode, though in general the features across the modes reported in Figs. 10 and 11 are all physically distinct. In contrast to similar attempts to visualize physical representations of individual latent directions (e.g., in [14]), the decodes are all smooth and physically interpretable, even at high Re.

B. Low-dimensional visualizations

The sparseness of the representation in the latent SVD basis motivates a lower-dimensional visualization. We present two-dimensional representations of the modal projections of the embeddings (12) for each snapshot in the test dataset using the t-SNE algorithm [58] for $\mathrm{Re} = 40$, $\mathrm{Re} = 100$, and $\mathrm{Re} = 400$. As local pairwise distances between snapshots in the (truncated) latent space are "preserved" in the t-SNE visualizations [58], nearby snapshots in these visualizations reflect nearby snapshots in the truncated latent space. When constructing these visualizations, we also include previously computed libraries of unstable periodic orbits (for full details see Refs. [14,15,39]) and ten short trajectories of 100 advective time units sampled at intervals of 0.1 time units when generating the t-SNE visualization. These snapshots had to be included when running the t-SNE algorithm, as it is a nonparametric self-supervised algorithm, but the unstable periodic orbits (UPOs) and trajectory representations are only plotted in the low-dimensional visualizations when explicitly stated in the text. The two-dimensional visualization plane is divided into 25×25 bins, and the mean dissipation rate of the snapshots in each bin is computed. Each point in the representation represents a snapshot in the dataset and is colored by this local mean dissipation rate. The contours in Figs. 12 and 14 indicate level curves of the local density of the visualization, computed via Gaussian kernels [59]. The qualitative features (the clusters) of the output are robust to different initializations of the algorithm.

The two-dimensional visualizations in Figs. 12 and 14 should be compared to those computed previously in PHBK and [35]; these were dominated by the learned shift-reflect symmetry, and the data appeared to form a large octagon with little other structure. The full symmetry reduction here, combined with the IRMAE approach, results in many more t-SNE clusters with distinct embedding features, each of which represents a distinctive class of vorticity snapshots in physical space.

At $\mathrm{Re} = 40$, representative flow states are visualized in Fig. 12 from a selection of low-dissipation t-SNE

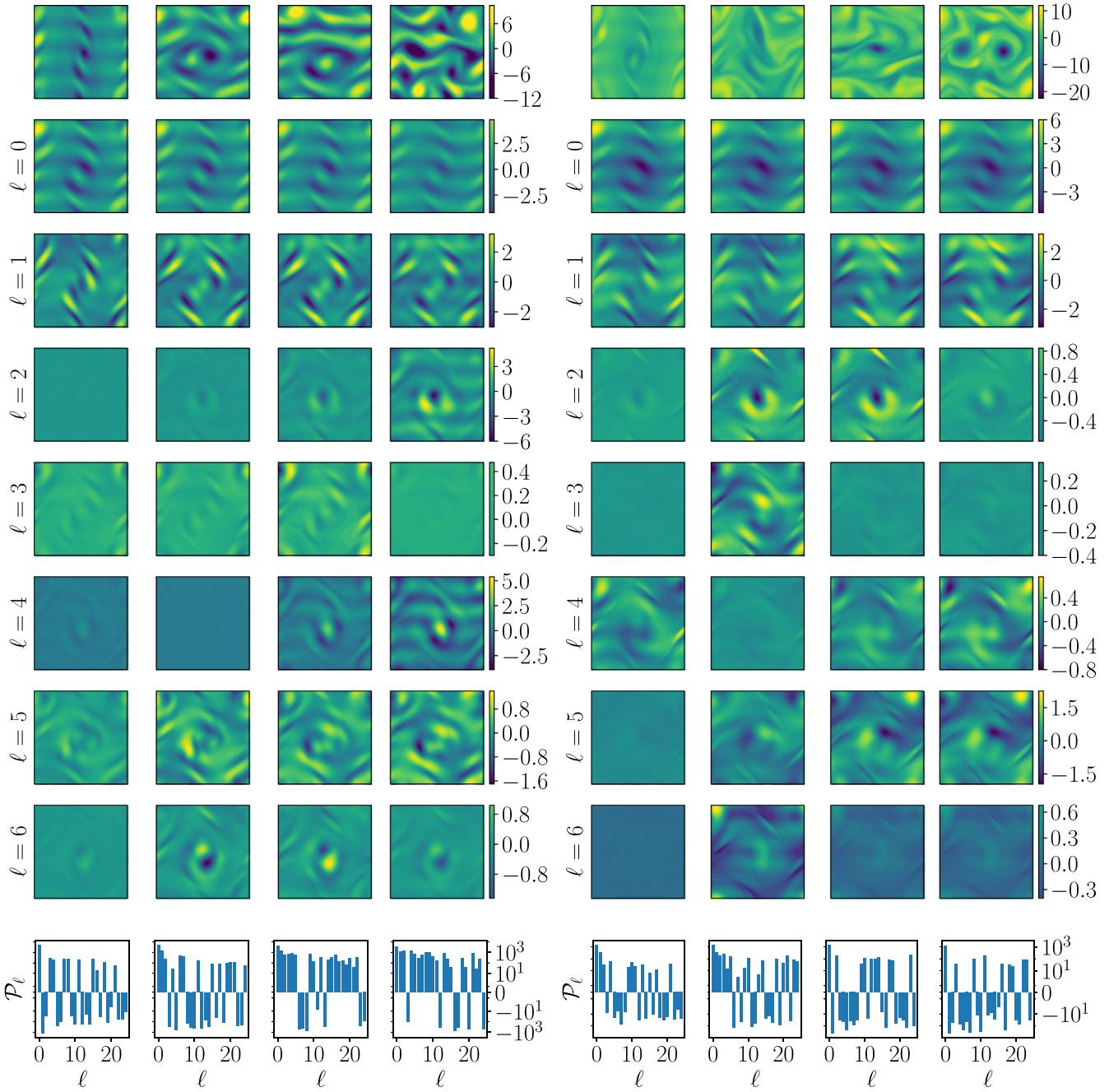


FIG. 10. The contributions $\mathcal{D}(\mathcal{P}_0(\omega)\mathbf{u}_0)$ from mode 0 and $\mathcal{D}(\mathcal{P}_\ell(\omega)\mathbf{u}_\ell + \mathcal{P}_0(\omega)\mathbf{u}_0) - \mathcal{D}(\mathcal{P}_0(\omega)\mathbf{u}_0)$ from modes $\ell = 1, \ldots, 6$. The top row shows the input vorticity snapshots at (left panels) Re $= 40$ with $D/D_{\text{lam}} = 0.06, 0.12, 0.18, 0.26$ from left to right, (right panels) Re $= 100$ with $D/D_{\text{lam}} = 0.016, 0.032, 0.047, 0.064$ from left to right. Each column shows the contribution from mode ℓ to the reconstruction of the input vorticity snapshot. Each snapshot in a row shares the same colorbar to highlight the activations of different modes across dissipation rates. The bottom row shows the projections $\mathcal{P}_\ell$ (on a logarithmic scale) onto the first 25 latent modes for the snapshot in each column.

clusters, with their respective positions in the t-SNE visualization indicated by arrows. The clusters all show the characteristic zigzag "staircase" patterns in the vorticity and are distinguished by the number and relative locations of dominant vortices. Two clusters on the left of the t-SNE visualization are notably detached from the bulk of the visualization. They sit far from the higher dissipation events and contain the lowest dissipation fields with large patches of quiescent fluid (see the representative snapshots with a white circle in the top left corner). The clear clustering into many distinct classes of events, not visible in previous work [14,35], and apparent distinction of different types of high-dissipation dynamics motivates further investigation into the nature of bursting events and routes into/out of this region of state space.

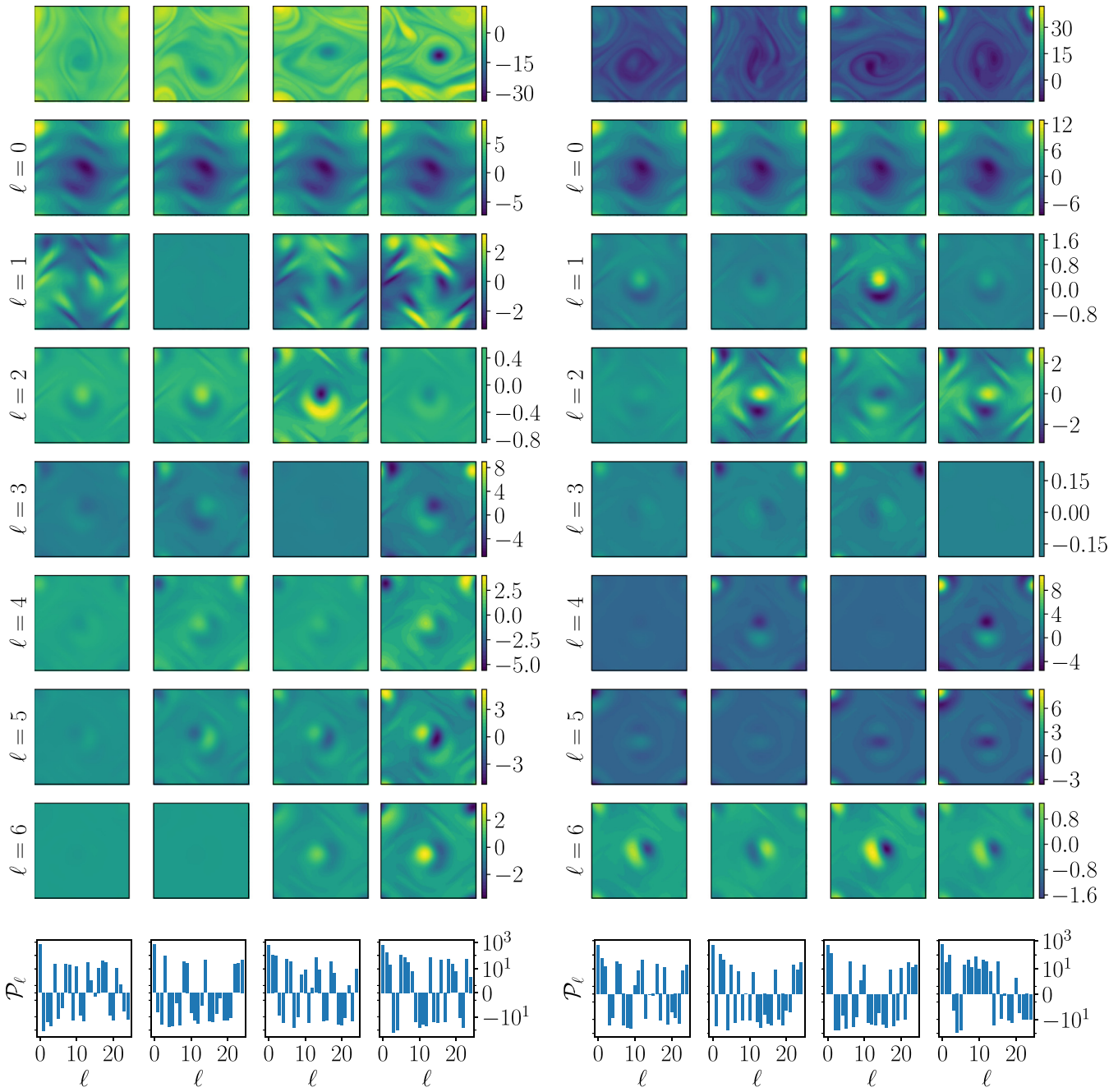


FIG. 11. The contributions $\mathcal{D}(\mathcal{P}_0(\omega)\mathbf{u}_0)$ from mode 0 and $\mathcal{D}(\mathcal{P}_\ell(\omega)\mathbf{u}_\ell + \mathcal{P}_0(\omega)\mathbf{u}_0) - \mathcal{D}(\mathcal{P}_0(\omega)\mathbf{u}_0)$ from modes $\ell = 1, \ldots, 6$. The top row shows the input vorticity snapshots at (left panels) Re $= 200$ with $D/D_{\text{lam}} = 0.0064, 0.0088, 0.0102, 0.0233$ from left to right, (right panels) Re $= 400$ with $D/D_{\text{lam}} = 0.0022, 0.003, 0.0035, 0.0054$ from left to right. Each column shows the contribution from mode ℓ to the reconstruction of the input vorticity snapshot. Each snapshot in a row shares the same colorbar to highlight the activations of different modes across dissipation rates. The bottom row shows the projections $\mathcal{P}_\ell$ (on a logarithmic scale) onto the first 25 latent modes for the snapshot in each column.

1. Bursting classes

The high-dissipation or "bursting" events (yellow/orange/red markers in Fig. 12) are found in two (possibly three) distinct classes, one surrounded by lower-dissipation events, labeled 1 in Fig. 12, and another which sits apart in the visualization and is more diffuse, labeled 2. This second, higher-dissipation region may itself be made up of two subclasses, though the amount of data in this region makes this difficult to discern. The two classes of bursting events can be distinguished by a subset of latent SVD modes $\ell \in L_d$. This is done in the right panel of Fig. 12, where the most significant modes are identified by computing the mean projection onto each ℓ within the class, $\langle \mathcal{P}_\ell(\omega) \rangle_i$ ($i = 1$ or 2) (the average was taken over all snapshots satisfying $D/D_{\text{lam}} > D_c$ within manually defined boundaries on the t-SNE plane). The subset of modes with the largest difference

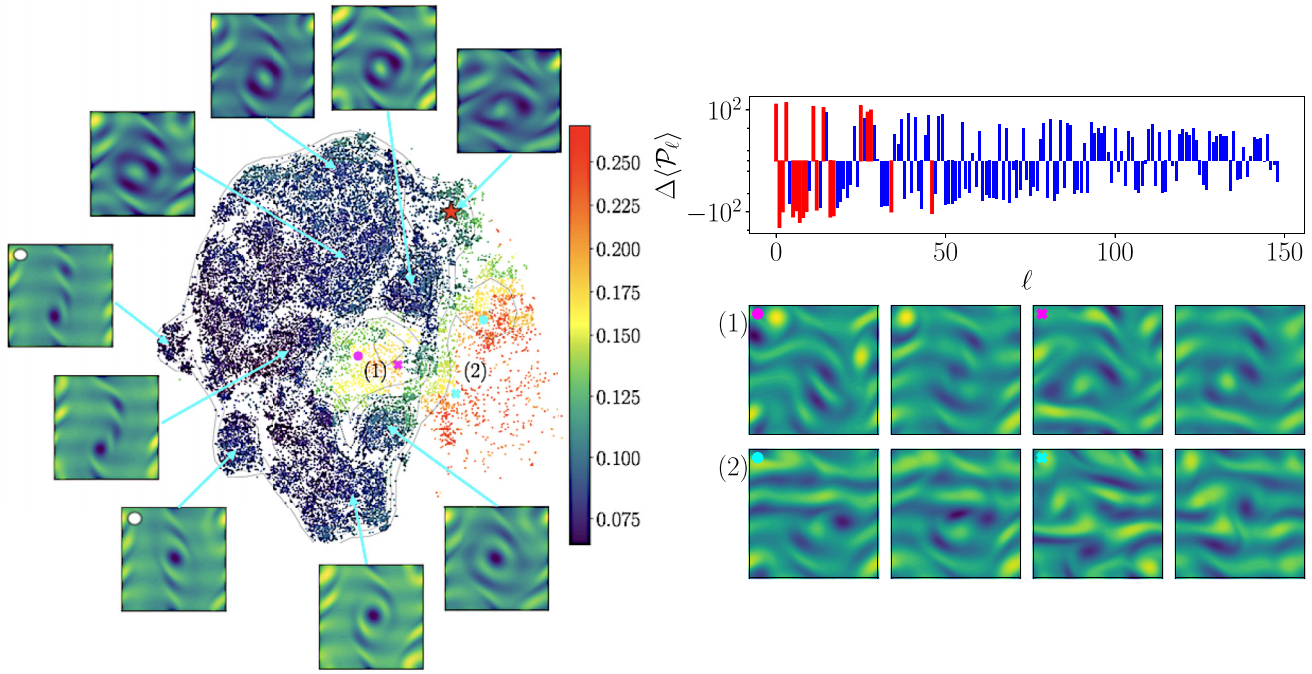


FIG. 12. Two-dimensional visualization of the coordinate vector (12) at Re = 40 using the t-SNE algorithm [58]. (Left) The x and y axes represent the two degrees of freedom in the t-SNE visualization, which are determined such that the similarity between neighboring mapped data (measured by t-Student distributions) approximates the similarity in the original variables (measured by Gaussian distributions). Each snapshot is colored by a binned mean normalized dissipation rate D/D_{lam}, where $D_{\text{lam}} = \text{Re}/(2n^2)$ is the laminar dissipation rate, using 25 bins in both x and y directions. The dotted black lines indicate level curves of the Gaussian kernel density estimates of the representation such that increasingly bold lines indicate increasing local density. The spanwise vorticity of representative states indicated by the arrows visualize the distinct dynamical processes contained within the high-density clusters of the visualization. Snapshots are extracted belonging to the bursting event class $i = 1, 2$, and the average projection onto each mode $\langle\mathcal{P}_\ell\rangle_i$ of class i is computed. (Right top) The difference $\Delta\langle\mathcal{P}_\ell\rangle := \langle\mathcal{P}_\ell\rangle_1 - \langle\mathcal{P}_\ell\rangle_2$ is shown for each mode ℓ. A subset of modes $\ell \in L_d$ (red bars) appears to be mostly responsible for delineating the two bursting event classes. (Right bottom) Two representative snapshots ω indicated by fuchsia (top panels) and cyan (bottom panels) markers from the two classes of bursting events are shown. The decodes $\mathcal{D}(\sum_{\ell\in L_d}\mathcal{P}_\ell(\omega)\mathbf{u}_\ell)$ are shown directly to the right of each representative snapshot ω.

$\Delta\langle\mathcal{P}_\ell\rangle := \langle\mathcal{P}_\ell\rangle_1 - \langle\mathcal{P}_\ell\rangle_2$ is highlighted in red in the right panel of Fig. 12; this defines the subset L_d.

The physical structures spanned by these modes are revealed in the bottom right panel of Fig. 12 by visualizing the decode of the sum of the projection onto these modes $\mathcal{D}(\sum_{\ell\in L_d}\mathcal{P}_\ell(\omega_i)\mathbf{u}_\ell)$ for representative snapshots ω_i from the two bursting classes which are indicated in the t-SNE plot. Class 1 (the calmer high-dissipation cluster) has a stronger contribution from the SVD mode 0, which was associated with the time-averaged mean flow (see Fig. 10). The signature of this large-scale mode is clearly visible in the decodes of the subset L_d in the bottom right panels of Fig. 12. In contrast, there is a weaker projection onto this mode for class 2, the higher dissipation class, which explains its location away from the bulk of the visualization in Fig. 12. Class 2 instead has a stronger projection onto more of the modes $\ell \in L_d$ and also excites more large-ℓ modes which describe smaller-scale vorticity features.

2. *Bursting trajectories*

Plotting highly temporally resolved trajectories in the t-SNE visualizations can be used to identify routes into and out of the high-dissipation regions, i.e., bursting events. This allows us to probe whether either bursting class is visited in isolation and where likely bursts originate from in state space. Within the t-SNE dataset are ten trajectories of length 100 advective units, and from these we extract and visualize all of the subtrajectories which either lead to bursting events from quiescent events, or vice versa. These subtrajectories are defined as the continuous sequences of snapshots satisfying $D < D_c$ that precede or follow the local dissipation maximum, where D_c is the same dissipation rate threshold used for the dataset resampling in Sec. II. We also require that the subtrajectories start from or finish at a sufficiently quiescent state satisfying $D < D_l$ and that the local dissipation maximum satisfies $D > D_h$, where $D_l, D_h = 0.125D_{\text{lam}}, 0.175D_{\text{lam}}$ for Re = 40. The dissipation rate and low-dimensional representation of one such bursting subtrajectory is shown in the middle panel of Fig. 13, where the dissipation rates of the prebursting and postbursting snapshots are indicated by the opaque markers.

Histograms of all pre- and postbursting events, visualized over the earlier t-SNE map of the state space, are reported in the top left and right panels of Fig. 13, respectively. There is a clear distinction between pre- and postbursting events. The plot reveals that the "calmer" bursting class 1 is not visited in isolation but is the route of

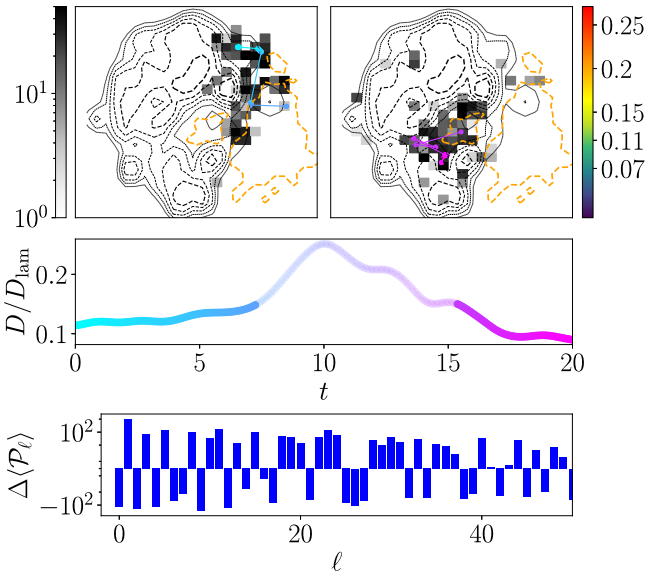


FIG. 13. (Top) Two-dimensional visualizations of the coordinate vector (12) at Re = 40 using the t-SNE algorithm [58]. The dotted black lines indicate level curves of the Gaussian kernel density estimates of the representation, with increasingly bold lines indicating increasing local density. The dotted orange line identifies the level curve of $D/D_{\text{lam}} = 0.15$. The grayscale histogram indicates the distribution of subtrajectories as defined in the text (left) which precede and (right) follow a bursting event. (Middle) The dissipation of a sample subtrajectory centered on the bursting event is plotted against time in advective time units, colored according to time. The opaque markers denote the pre- and postbursting subtrajectories extracted, while the transparent markers indicate the neglected bursting event. The same subtrajectories are shown by the connected, similarly colored points in the t-SNE visualization, demonstrating these sample routes (top left) toward and (top right) away from the bursting event. (Bottom) The difference $\Delta\langle\mathcal{P}_\ell\rangle := \langle\mathcal{P}_\ell\rangle_b - \langle\mathcal{P}_\ell\rangle_{\text{turb}}$ is shown for each mode ℓ, where $\langle\mathcal{P}_\ell\rangle_b$ and $\langle\mathcal{P}_\ell\rangle_{\text{turb}}$ denote the mean projection of snapshots belonging to the prebursting cluster (red star in Fig. 12) and all snapshots in the test dataset, respectively.

all high-dissipation events back to quiescent dynamics. This is confirmed by inspecting the full trajectories of bursting events on the t-SNE visualization (not shown), which always first hop between the two clusters within the extreme bursting class 2 before moving to the calmer bursting class 1.

The prebursting trajectories always visit the t-SNE cluster marked by a red star in Fig. 12, indicating that this cluster encapsulates some important prebursting vorticity features. To identify the associated distinguishing latent modes, the mean projection onto each mode ℓ for the snapshots within this cluster $\langle\mathcal{P}_\ell(\omega)\rangle_b$ and for all the snapshots in the test dataset sampled from the turbulent attractor $\langle\mathcal{P}_\ell(\omega)\rangle_{\text{turb}}$ are computed. The differences between these average projections are visualized in the lower panel of Fig. 13. This computation indicates that the prebursting cluster has a weaker projection onto the $\ell = 0$ mode representative of typical low-dissipation states. However, a much stronger projection onto the $\ell = 1$ mode, which features staggered diagonal stripes of vorticity (see Fig. 10) similar to unstable directions associated with many exact

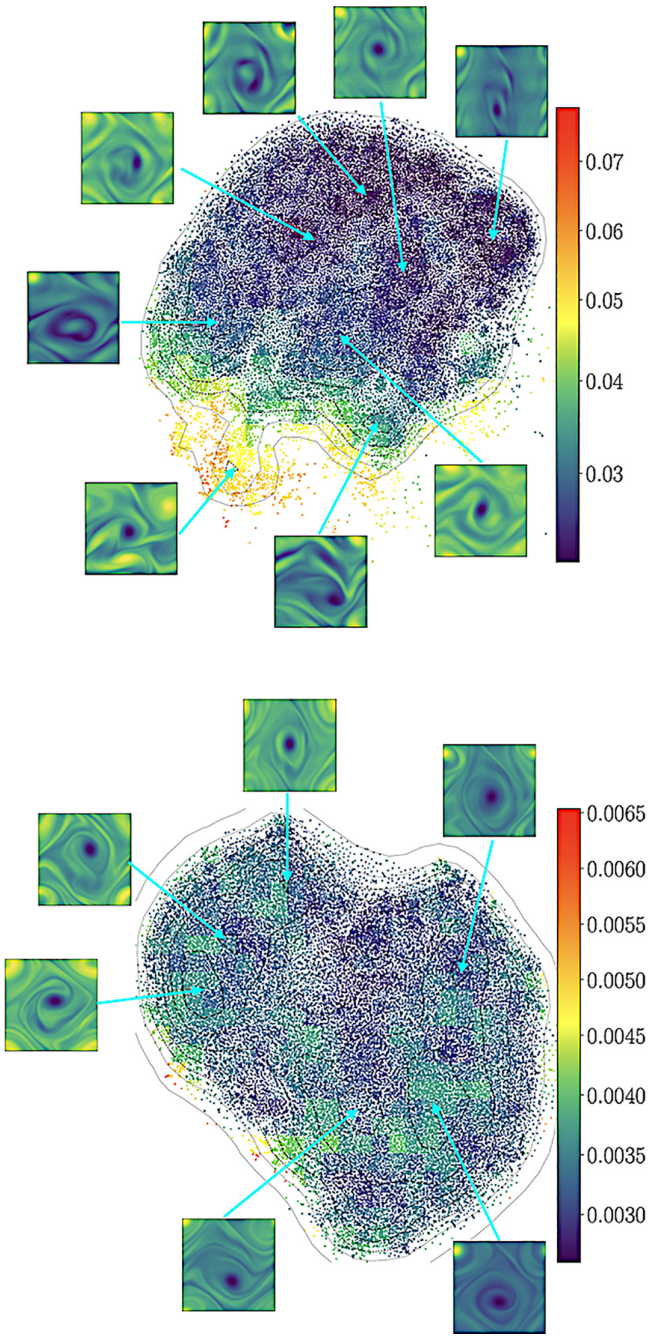


FIG. 14. Two-dimensional visualizations of the coordinate vector (12) at (top) Re = 100 and (bottom) Re = 400 using the t-SNE algorithm [58], as in Fig. 12, with visualizations of representative snapshots from the dense clusters.

solutions that resemble the mean flow state (see Fig. 18), is the distinguishing feature of the prebursting dynamical process within this cluster.

We have also confirmed that prebursting and postbursting regions on the visualization are both characterized by small $|\hat{\omega}(1,0)|$, while the bursting regions have a large value for $|\hat{\omega}(0,n)|$ (not shown). This large value in the bursting region is to be expected, as the production rate of Kolmogorov flow can be expressed by $I = -\hat{\omega}_R(0,n)/n$. The small value of $|\hat{\omega}(1,0)|$ in the prebursting region and the large value of

$|\hat{\omega}(0, n)|$ in the bursting region confirms the significant redistribution of energy between these modes before a bursting event, an observation which was also made in the variational approach of [42].

3. Higher Re *visualizations*

At Re = 100 (top panel of Fig. 14), many high-dissipation snapshots are spaced across the lower periphery of the two-dimensional visualization. There is also a distinct detached high-dissipation cluster indicating a distinct class of events which could not be separated cleanly in earlier models [14,35]. Although the clustering of data is less obvious than at the lower Re = 40 value, there are still distinct classes of vorticity snapshots that can be identified in this visualization. Some representative snapshots sampled from these clusters are shown in the top panel of Fig. 14. This figure indicates that the classes can be delineated based on the appearance of the large-scale vortices in the flow—their number, relative location, and strength. The sample from the detached high-dissipation cluster (bottom left of the visualization) again indicates an event where the importance of "mean flow" modes (e.g., $\ell = 0$ and 1) is diminished.

Finally, the t-SNE visualization of the most turbulent flow considered at Re = 400 (bottom panel of Fig. 14) no longer appears to cluster events based on a dominant dissipation rate. The high-dissipation snapshots are distributed fairly uniformly in the latent t-SNE visualization. The exception here are a couple of small pockets of very high-dissipation snapshots at the left periphery of the visualization. This behavior is consistent with the prominence of the large vortex pair associated with the condensate, a state which is apparent in most snapshots and hence leads to a more homogeneous latent representation across the dataset. Some representative snapshots of the high-density regions of the t-SNE visualization are again presented in the bottom panel of Fig. 14, and all show a large-scale vortex pair. Distinctions between snapshots are likely based on smaller-scale features (e.g. vortex filaments) and the relative location and/or shape of the dominant vortices.

C. Periodic orbit coverage

The motion of a turbulent flow within the turbulent attractor is postulated to be shaped and guided by unstable periodic orbits (UPOs)—exact solutions of the Navier–Stokes equations that each encapsulate some closed cycle of physical processes relevant to the turbulence itself [60–62]. A key motivation for this viewpoint is the possibility to connect dynamical events in the flow to statistical properties of turbulence, e.g., the cascade(s), via periodic orbit theory [63,64], which provides weights for expansions of the statistics of a chaotic attractor in terms of the statistics of the UPOs.

Large numbers of solutions have been amassed recently for the Kolmogorov flow studied here [14,15,39] at Re $\in \{40, 100\}$. These solutions are visualized in Fig. 15 in terms of their production and dissipation rates, overlayed on the turbulent distribution. These low-dimensional visualizations, and others like them, would seem to indicate the dynamical relevance [65] of the solutions due to their apparent proximity to the turbulence. However, these visualizations are not nuanced enough to identify the range of events contained in the library of UPOs and may give a misleading impression as to the relevance of a given solution to the true turbulent dynamics. The t-SNE plots constructed above provide a way to visualize the degree to which the range of true vortical events in the flow is covered by the library of solutions. As nearby snapshots in the t-SNE visualizations reflect similar snapshots in the truncated latent space, the visualizations effectively reveal whether the UPOs are similar to samples from the chaotic attractor and provide an alternative way to assess dynamical relevance.

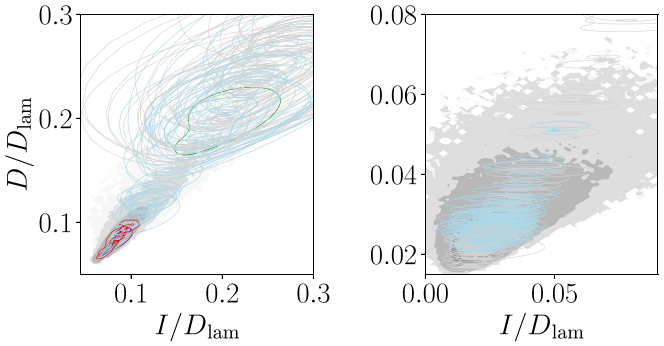


FIG. 15. Dissipation rate D against production rate I of the UPOs in the libraries of (left) 174 solutions at Re = 40 and (right) 252 solutions at Re = 100, both normalized by the laminar dissipation $D_{\text{lam}} = \text{Re}/(2n^2)$. The gray background is the probability density function (PDF) computed from a trajectory of 10^5 samples, separated by 1 advective time unit. The contour levels of the PDF are spaced logarithmically. Solutions deemed "dynamically relevant" by the simple dissipation-energy comparison described in the text are colored in light blue, while those not satisfying the condition are colored in gray.

1. Re = 40 *UPOs*

The Re = 40 UPOs are visualized within the two-dimensional t-SNE visualization in Fig. 16. The solutions are colored gray or blue according to a simple *a priori* dynamical relevance criteria of whether the time-averaged dissipation rate $\langle D \rangle$ and energy $\langle E \rangle$ lie within the first and ninety-ninth quantiles of the corresponding turbulent distributions. While some high-density regions of the turbulence feature snapshots from the UPOs, there are large areas of high turbulent-snapshot density which do not contain any at all. This is surprising given the robustness of the statistical reconstructions possible with this set of solutions observed in Refs. [15,39].

When only the shorter orbits (here $T < 5$) are visualized (see top right panel of Fig. 16), the picture simplifies further. The low-dissipation, short orbits are confined to the edges of the visualization. On the other hand, there are many high-dissipation UPOs which were previously labeled "irrelevant" (based on E and D – gray symbols) that do overlap with the t-SNE embeddings of turbulent snapshots. This indicates that turbulent orbits may visit the neighborhoods of these solutions despite the mismatch in their production and dissipation values. Snapshots from some individual, modest-length UPOs are also included in Fig. 16 and are identified in the

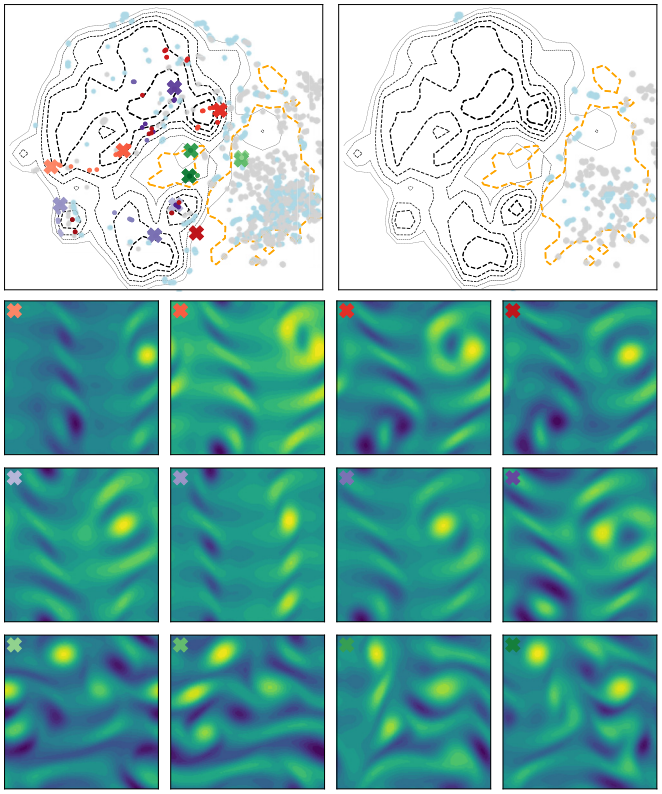


FIG. 16. (Top) Two-dimensional visualizations of the coordinate vector (12) at Re = 40 using the t-SNE algorithm [58]. The dotted black lines indicate level curves of the Gaussian kernel density estimates of the representation such that increasingly bold lines indicate increasing local density. The dotted orange line indicates the level curve of $D/D_{\text{lam}} = 0.15$. The points indicate snapshots sampled along the periods of the "dynamically relevant" (light blue) and "dynamically irrelevant" (gray) UPOs in (left) the full library of 174 solutions and (right) the short-period orbits with periods $T < 5$ advective time units. Fifty snapshots are sampled for each UPO. (Bottom) Each row shows equally spaced snapshots (without full symmetry preprocessing) along the periods of three periodic orbits with periods $T = 23.16$, 18.77, and 5.06, indicated in the t-SNE visualization by the red, purple, and green points, respectively, from top to bottom, where increasing boldness denotes time evolution along the period of the orbit. The crosses in the right top panel indicate the positions of the corresponding visualized snapshots.

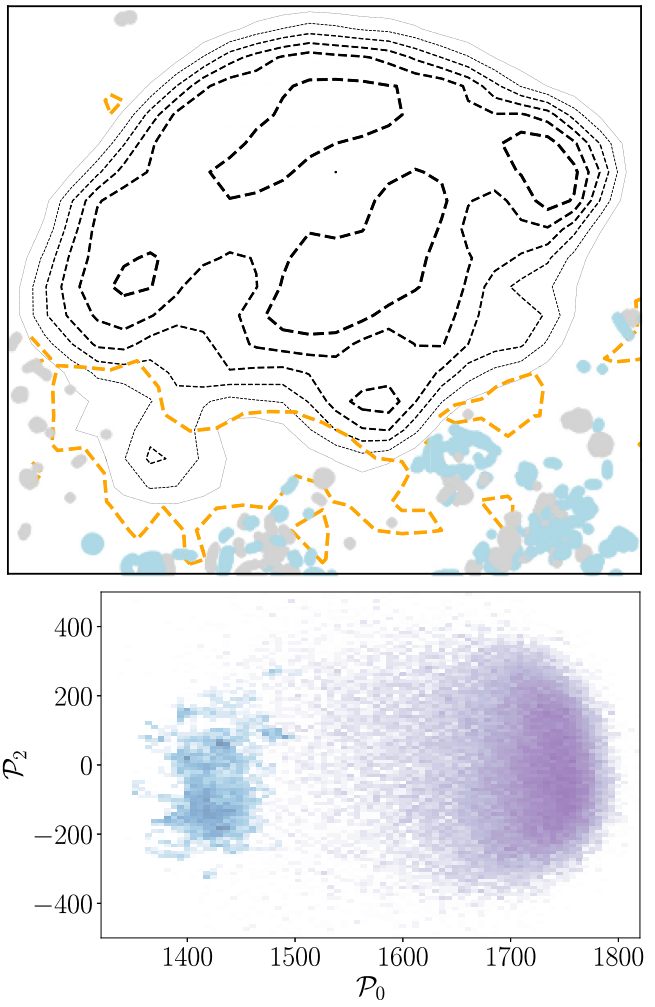


FIG. 17. (Top) Two-dimensional visualizations of the coordinate vector (12) at Re = 100 using the t-SNE algorithm [58]. The dotted black lines indicate level curves of the Gaussian kernel density estimates of the representation such that increasingly bold lines indicate increasing local density. The dotted orange line indicates the level curve of $D/D_{\text{lam}} = 0.045$. The points indicate snapshots sampled along the periods of the "dynamically relevant" (light blue) and "dynamically irrelevant" (gray) UPOs in the full library of 252 solutions. Fifty snapshots are sampled for each UPO. (Bottom) The joint distribution of the latent projections $\mathcal{P}_0$ and $\mathcal{P}_2$ of snapshots sampled from the periods of the "dynamically relevant" UPOs (blue) and snapshots sampled from the turbulent attractor (purple). The densities are colored on a log scale.

t-SNE visualization. These UPOs are also colored accordingly in the dissipation versus production rate diagram in the left panel of Fig. 15. The two longest period orbits, with period $T = 23.16$ (red) and $T = 18.77$ (purple), visit many of the high-density clusters throughout the visualization. In contrast, the $T = 5.06$ (green) orbit is shorter and more localized to a high-dissipation region of the visualization. This UPO is the only solution in the library that visits the smaller bursting cluster at Re = 40 discussed in Sec. V B.

2. Re = 100 *UPOs*

A similar visualization is conducted for the UPOs computed at Re = 100 in the top panel of Fig. 17. While the vast majority of UPOs have production and dissipation values consistent with that of the turbulence (see Fig. 15), the t-SNE map indicates that nearly *all* UPOs are determined by the autoencoder to be distinct from the turbulent dataset. This is quite striking given the overlap in various diagnostic statistics reported here and elsewhere [15,38,39]. The apparent lack of similarity can be diagnosed by considering the projections $\mathcal{P}_\ell(\omega)$ onto individual latent directions.

To investigate the cause for the separation of these UPOs from the turbulent attractor, the mean projection onto each mode ℓ was computed for all the snapshots sampled from the "dynamically relevant" (as determined naively from the dissipation and energy PDFs) UPOs at Re = 100, $\langle \mathcal{P}_\ell(\omega) \rangle_{\text{UPO}}$, and for all snapshots in the test dataset sampled from the turbulent

attractor at $\mathrm{Re} = 100$, $\langle \mathcal{P}_\ell(\omega) \rangle_{\mathrm{turb}}$. Differences were found in the distributions for large numbers of the latent projections, but the most significant distinction between the turbulence and the UPOs was found in the leading mode $\ell = 0$. Joint distributions for both the UPO data and turbulence are reported in the bottom panel of Fig. 17 (the other mode is selected as $\ell = 2$ here). This figure highlights the substantial difference in the $\ell = 0$ distributions; the UPO or turbulence data overlap only slightly in the tails of the distributions. Recall that in Sec. V the $\ell = 0$ mode was decoded and shown to be visually similar to the symmetry-reduced, time-average flow state in the turbulence. The dramatic reduction of importance of this mode should also be considered in the context of the continuation study reported in [39], which showed that many $\mathrm{Re} = 100$ solutions appear to "move away" from the turbulent attractor as Re increases. The results reported here would suggest that perhaps they are not inside the turbulent attractor, but just "close by" at $\mathrm{Re} = 100$, consistent with the continuation results.

VI. CONCLUSION

In this study we designed deep autoencoders for learning low-dimensional representations of 2D Kolmogorov flow. The models combined (i) full discrete and continuous symmetry reduction, (ii) dense convolutional input-output streams, and (iii) an "implicit rank minimizing" (IRMAE [28,32]) component. The autoencoders were used to estimate the dimensionality of the turbulent attractor in the two-dimensional Navier–Stokes equations, which was obtained from the rank of the innermost network representation. An approximate scaling of $\tilde{d}_{\mathcal{A}} \sim \mathrm{Re}^{1/3}$ was found in the fully turbulent regime, a much weaker scaling than the known upper and lower bounds on the *global* attractor [8–10].

The combination of the IRMAE approach with a full symmetry reduction yielded a rich latent-space structure with which individual latent SVD modes could be decoded into physically interpretable structures at all Re values.

Two-dimensional visualizations of large turbulent datasets revealed distinct clusters corresponding to different classes of vorticity snapshots. At $\mathrm{Re} = 40$, multiple types of high-dissipation events were identified. Visualizations of finely resolved bursting trajectories highlighted a common route to a bursting event involving a large projection onto a particular dominant latent mode and a common route back to quiescent flow through one of the high-dissipation clusters.

The low-dimensional visualizations of turbulence were then used to assess the coverage of the turbulent dynamics by large libraries of unstable periodic orbits at both $\mathrm{Re} = 40$ and $\mathrm{Re} = 100$. There was a reasonable representation, with notable gaps, at the lower Re value. However, the visualizations at $\mathrm{Re} = 100$ revealed that the latent representations of nearly *all* periodic solutions were distinct from the turbulence in a fundamental way. An analysis of individual latent structures indicated a much weaker projection onto a structure associated with the turbulent mean profile. These distinctions are not apparent in simpler low-dimensional projections (e.g., production/dissipation) and indicate that these solutions may not be embedded in the turbulent attractor.

Autoencoders seek to exactly reconstruct their inputs, and hence many of the $O(\mathrm{Re}^{1/3})$ dimensions found here may not correspond directly to dynamical variables. For instance, it is known in three-dimensional turbulence that scales smaller than $\sim 5\pi\eta$ (where η is the Kolmogorov length) are "slaved" to the larger scales within the turbulent attractor [66,67]. The equivalent result in two-dimensional turbulence is unknown at present, but having a model learn only the "dynamical" components of vorticity, determined via an appropriate low-pass filter, would presumably simplify the reconstruction problem and allow for a more robust estimate of the Re dependence. We hope to report on this in the near future.

ACKNOWLEDGMENTS

This research has been supported by the UK Engineering and Physical Sciences Research Council through the MAC-MIGS Centre for Doctoral Training (EP/S023291/1). Computations were performed on the Cirrus UK National Tier-2 HPC Service at EPCC [68]. J.P. acknowledges support from UKRI Frontier Guarantee Grant No. EP/Y004094/1. We are grateful to Masanobu Inubushi for pointing us to Ref. [10] with a lower bound for the dimension of the global attractor. J.P. also thanks Rich Kerswell for useful discussions and Mike Graham for helpful discussion on the IRMAE method.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request. The full source code is available at [45].

APPENDIX A: ARCHITECTURAL DETAILS

Full details of the architecture used in this work are presented in this section. Dense blocks [43,44] are used here in place of traditional single convolutional layers in both the encoder and decoder, similar to the approach in Ref. [14]. The blocks are groups of convolutional layers where the output of each convolutional layer is concatenated with its input. Each dense block consists of three convolutional layers with 32 filters each. Periodic padding is also applied to each image before a convolution, so that the output of a convolutional layer has the same shape as the input. The GELU activation function [69] and batch normalization [70] are applied to the output of convolutional layers. Max pooling layers are used to gradually reduce the dimensionality of the image throughout the encoder, while upsampling layers are applied within the decoder to return the original shape at the decoder output.

The encoder architecture can be summarized by the following sequence of operations:

$$\begin{aligned}\omega &\to \mathrm{PC}(8\times 8, 32) \to \mathrm{DB}(8\times 8) \to \mathrm{MP}(2\times 2)\\ &\to \mathrm{PC}(4\times 4, 32) \to \mathrm{DB}(4\times 4) \to \mathrm{MP}(2\times 2)\\ &\to \mathrm{PC}(4\times 4, 32) \to \mathrm{DB}(4\times 4) \to \mathrm{MP}(2\times 2)\\ &\to \mathrm{PC}(4\times 4, 32) \to \mathrm{DB}(4\times 4)\end{aligned}$$

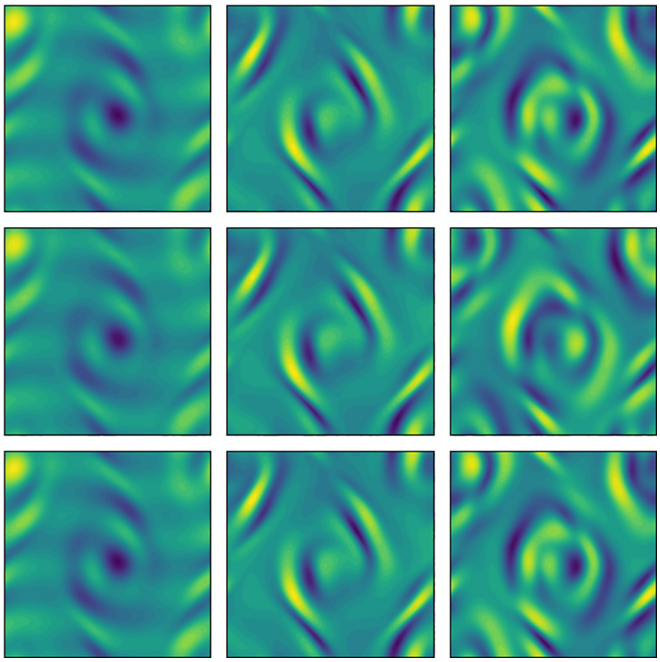


FIG. 18. Out-of-plane vorticity for three low-dissipation exact solutions at Re = 40 (left), alongside the real (middle) and imaginary (right) components of the leading unstable eigenfunction. (Top) A traveling wave solution with a wavespeed of 0.0198. (Middle) A UPO with period $T = 12.207$ and translational shift of $s = 0.243$. (Bottom) A UPO with $T = 5.81$ and $s = 0.115$.

$$\rightarrow \mathrm{PC}(4\times 4, 4) \rightarrow \mathrm{Flatten} = \mathscr{E}(\omega), \tag{A1}$$

where the first term in the brackets denotes the size of the convolutional filters and the second term denotes the number of filters, PC stands for a convolutional layer with periodic padding, DB for the dense blocks described above, and MP for a max pooling layer. The input vorticity field is a single-channel image of size 128^2, and hence the output of the encoder $\mathscr{E}(\omega)$ is a flattened vector of length $16^2 \times 4 = 1024$.

The vector $\mathscr{E}(\omega)$ is then passed through a series of four fully connected (FC) linear layers with equal input-output dimension. These layers do not include bias terms and hence are pure matrix multiplications of the upstream input. The architecture of the bottleneck is thus summarized by

$$\mathscr{E}(\omega) \rightarrow FC(1024)^4 = \mathscr{W}(\mathscr{E}(\omega)). \tag{A2}$$

This output $\mathscr{W}(\mathscr{E}(\omega))$ is the latent representation learned by the autoencoder. The impact of the internal linear operations is to drive the representation to low rank, something which has also been empirically observed in other studies [28,32].

To decode this latent representation back to physical space, $\mathscr{W}(\mathscr{E}(\omega))$ is reshaped into an image with four channels of dimension 16^2. The structure of the decoder used then essentially mirrors the encoder (replacing max pooling layers with upsampling layers (UP)), with the exception of the addition of an output stream of linear periodic convolutional layers (LPC):

$$\begin{aligned}\mathscr{W}(\mathscr{E}(\omega)) &\rightarrow \mathrm{Reshape}\\ &\rightarrow \mathrm{PC}(4\times 4, 32) \rightarrow \mathrm{DB}(4\times 4) \rightarrow \mathrm{UP}(2\times 2)\\ &\rightarrow \mathrm{PC}(4\times 4, 32) \rightarrow \mathrm{DB}(4\times 4) \rightarrow \mathrm{UP}(2\times 2)\\ &\rightarrow \mathrm{PC}(8\times 8, 32) \rightarrow \mathrm{DB}(8\times 8)\\ &\rightarrow \mathrm{LPC}(16\times 16, 1) \rightarrow \mathrm{LPC}(12\times 12, 1)\\ &\rightarrow \mathrm{LPC}(8\times 8, 1) \rightarrow \mathrm{LPC}(6\times 6, 1)\\ &= \mathscr{D}[\mathscr{W}(\mathscr{E}(\omega))] = \mathscr{A}(\omega).\end{aligned} \tag{A3}$$

These linear layers result in a more faithful reproduction of the spectral content of the input snapshots (see Sec. III).

TABLE II. Sensitivity of the average ε and standard deviation σ of the relative error on the initial learning rate and batch size.

Initial learning rate	Batch size	ε	σ
10^{-4}	16	0.021	0.015
5×10^{-3}	16	0.032	0.0196
10^{-3}	16	0.038	0.025
10^{-4}	8	0.0186	0.0116
10^{-4}	16	0.0187	0.0141
10^{-4}	64	0.0188	0.0133

APPENDIX B: UNSTABLE EIGENFUNCTIONS

We include here three examples of low-dissipation exact solutions at Re = 40, along with a visualization of their leading unstable eigenfunctions. These snapshots are reported in Fig. 18 and should be compared with decodes of the leading SVD modes reported in Fig. 10.

APPENDIX C: HYPERPARAMETER SENSITIVITY

Hyperparameter sensitivity for the IRMAE networks was checked using a smaller (two convolutional layers per dense block, smaller embedding dimension of $16^2 = 256$) version of the final network outlined in Appendix A. A smaller network was used to perform this hyperparameter analysis as the full IRMAE networks required ∼2 weeks of wall-time due to the large datasets and small batch sizes considered. The average and standard deviation of the relative error are reported in

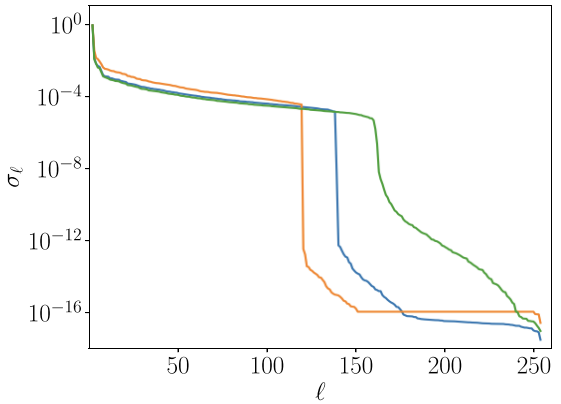


FIG. 19. Singular values σ_ℓ in the singular value decomposition of the covariance of the latent data matrix for a smaller network trained with a batch size of 64 (green), 16 (blue), and 8 (orange).

Table II for a number of combinations of batch size and initial learning rates.

An initial learning rate of 10^{-4} yields the best relative error, while the relative error is quite robust to changes in the batch size. However, the rank of the learned representation is significantly smaller for smaller batch sizes, as shown in Fig. 19. Keeping the size of the training dataset and the number of epochs constant, a smaller batch size implies that a greater number of gradient descent steps are taken. This result is supportive of the idea that each gradient update of the weights incrementally drives the rank of the learned latent space downward.

[1] E. Hopf, A mathematical example displaying features of turbulence, Commun. Pure Appl. Math. **1**, 303 (1948).

[2] R. Temam, *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*, Applied Mathematical Sciences Vol. 68 (Springer, New York, 1997).

[3] C. Foias, M. S. Jolly, I. G. Kevrekidis, G. R. Sell, and E. S. Titi, On the computation of inertial manifolds, Phys. Lett. A **131**, 433 (1988).

[4] E. S. Titi, On approximate inertial manifolds to the Navier-Stokes equations, J. Math. Anal. Appl. **149**, 540 (1990).

[5] C. R. Doering and J. D. Gibbon, *Applied Analysis of the Navier-Stokes Equations* (Cambridge University Press, Cambridge, England, 1995).

[6] R. Temam, Do inertial manifolds apply to turbulence? Physica D **37**, 146 (1989).

[7] C. Foias, O. Manley, and R. Temam, Modelling of the interaction of small and large eddies in two dimensional turbulent flows, ESAIM: Math. Modell. Numer. Anal. **22**, 93 (1988).

[8] P. Constantin, C. Foias, and R. Temam, On the dimension of the attractors in two-dimensional turbulence, Physica D **30**, 284 (1988).

[9] C. R. Doering and J. D. Gibbon, Note on the Constantin-Foias-Temam attractor dimension estimate for two-dimensional turbulence, Physica D **48**, 471 (1991).

[10] V. X. Liu, A sharp lower bound for the Hausdorff dimension of the global attractors of the 2D Navier-Stokes equations, Commun. Math. Phys. **158**, 327 (1993).

[11] R. H. Kraichnan, Inertial ranges in two-dimensional turbulence, Phys. Fluids **10**, 1417 (1967).

[12] R. H. Kraichnan, Inertial-range transfer in two- and three-dimensional turbulence, J. Fluid Mech. **47**, 525 (1971).

[13] K. Ohkitani, Log-corrected energy spectrum and dimension of attractor in two-dimensional turbulence, Phys. Fluids **1**, 451 (1989).

[14] J. Page, J. Holey, M. P. Brenner, and R. R. Kerswell, Exact coherent structures in two-dimensional turbulence identified with convolutional autoencoders, J. Fluid Mech. **991**, A10 (2024).

[15] J. Page, P. Norgaard, M. P. Brenner, and R. R. Kerswell, Recurrent flow patterns as a basis for two-dimensional turbulence: Predicting statistics from structures, Proc. Natl. Acad. Sci. USA **121**, e2320007121 (2024).

[16] J. L. Kaplan and J. A. Yorke, Chaotic behavior of multidimensional difference equations, in *Functional Differential Equations and Approximation of Fixed Points (Proceedings of the Summer School and Conference, University of Bonn, 1978)*, Lecture Notes in Mathematics Vol. 730 (Springer, Berlin, 1979), pp. 204–227.

[17] R. Grappin and J. Léorat, Computation of the dimension of two-dimensional turbulence, Phys. Rev. Lett. **59**, 1100 (1987).

[18] R. Grappin and J. Léorat, Lyapunov exponents and the dimension of periodic incompressible Navier–Stokes flows: Numerical measurements, J. Fluid Mech. **222**, 61 (1991).

[19] D. Clark, L. Tarra, and A. Berera, Chaos and information in two-dimensional turbulence, Phys. Rev. Fluids **5**, 064608 (2020).

[20] H. Zou, T. Hastie, and R. Tibshirani, Sparse principal component analysis, J. Comput. Graphical Stat. **15**, 265 (2006).

[21] C. Bishop, Bayesian PCA, in *Advances in Neural Information Processing Systems*, edited by M. Kearns, S. Solla, and D. Cohn (MIT Press, Cambridge, MA, 1998), Vol. 11.

[22] A. J. Linot and M. D. Graham, Deep learning to discover and predict dynamics on an inertial manifold, Phys. Rev. E **101**, 062209 (2020).

[23] A. J. Linot and M. D. Graham, Data-driven reduced-order modeling of spatiotemporal chaos with neural ordinary differential equations, Chaos **32**, 073110 (2022).

[24] P. R. Vlachas, G. Arampatzis, C. Uhler, and P. Koumoutsakos, Multiscale simulations of complex systems by learning their effective dynamics, Nat. Mach. Intell. **4**, 359 (2022).

[25] C. E. P. De Jesús and M. D. Graham, Data-driven low-dimensional dynamic model of Kolmogorov flow, Phys. Rev. Fluids **8**, 044402 (2023).

[26] A. J. Linot and M. D. Graham, Dynamics of a data-driven low-dimensional model of turbulent minimal Couette flow, J. Fluid Mech. **973**, A42 (2023).

[27] K. Zeng, C. E. P. De Jesús, A. J. Fox, and M. D. Graham, Autoencoders for discovering manifold dimension and coordinates in data from complex dynamical systems, Mach. Learn.: Sci. Technol. **5**, 025053 (2024).

[28] L. Jing, J. Zbontar, and Y. LeCun, Implicit rank-minimizing autoencoder, in *Advances in Neural Information Processing Systems*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (Curran Associates, Inc., Red Hook, New York, 2020), Vol. 33, pp. 14736–14746.

[29] C. E. P. D. Jesús, A. J. Linot, and M. D. Graham, Building symmetries into data-driven manifold dynamics models for complex flows: Application to two-dimensional Kolmogorov flow, arXiv:2312.10235.

[30] M. Y. Vinograd and P. Clark Di Leoni, Reduced representations of Rayleigh–Bénard flows via autoencoders, J. Fluid Mech. **1006**, A10 (2025).

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, in *Advances in Neural Information Processing Systems*, edited by F. Pereira, C. Burges, L. Bottou, and K. Weinberger (Curran Associates, Inc., Red Hook, NY, 2012), Vol. 25.

[32] K. Zeng and M. D. Graham, Symmetry reduction for deep reinforcement learning active control of chaotic spatiotemporal dynamics, Phys. Rev. E **104**, 014210 (2021).

[33] N. B. Budanur, D. Borrero-Echeverry, and P. Cvitanović, Periodic orbit analysis of a system with continuous symmetry—A tutorial, Chaos **25**, 073112 (2015).
[34] S. Kneer, T. Sayadi, D. Sipp, P. Schmid, and G. Rigas, Symmetry-aware autoencoders: s-PCA and s-nlPCA, arXiv:2111.02893.
[35] J. Page, M. P. Brenner, and R. R. Kerswell, Revealing the state space of turbulence using machine learning, Phys. Rev. Fluids **6**, 034402 (2021).
[36] A. Racca, N. A. K. Doan, and L. Magri, Predicting turbulent dynamics with the convolutional autoencoder echo state network, J. Fluid Mech. **975**, A2 (2023).
[37] Z. Y. Wan, P. Vlachas, P. Koumoutsakos, and T. Sapsis, Data-assisted reduced-order modeling of extreme events in complex dynamical systems, PLoS One **13**, e0197704 (2018).
[38] G. J. Chandler and R. R. Kerswell, Invariant recurrent solutions embedded in a turbulent two-dimensional Kolmogorov flow, J. Fluid Mech. **722**, 554 (2013).
[39] A. Cleary and J. Page, Dynamical relevance of periodic orbits under increasing Reynolds number and connections to inviscid dynamics, J. Fluid Mech. **1020**, A52 (2025).
[40] D. Kochkov, J. A. Smith, A. Alieva, Q. Wang, M. P. Brenner, and S. Hoyer, Machine learning–accelerated computational fluid dynamics, Proc. Natl. Acad. Sci. USA **118**, e2101784118 (2021).
[41] G. Dresdner, D. Kochkov, P. Norgaard, L. Zepeda-Núñez, J. A. Smith, M. P. Brenner, and S. Hoyer, Learning to correct spectral methods for simulating turbulent flows, arXiv:2207.00556.
[42] M. Farazmand and T. P. Sapsis, A variational approach to probing extreme events in turbulent dynamical systems, Sci. Adv. **3**, e1701533 (2017).
[43] G. Huang, Z. Liu, and K. Q. Weinberger, Densely connected convolutional networks, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, New York, 2016).
[44] G. Huang, Z. Liu, G. Pleiss, L. J. P. van der Maaten, and K. Q. Weinberger, Convolutional networks with dense connectivity, IEEE TPAMI **44**, 8704 (2019).
[45] https://github.com/computational-flow-science/IRMAE.
[46] A. Solera-Rico, C. Sanmiguel Vila, M. Gómez-López, Y. Wang, A. Almashjary, S. T. M. Dawson, and R. Vinuesa, β-variational autoencoders and transformers for reduced-order modelling of fluid flows, Nat. Commun. **15**, 1361 (2024).
[47] J. Li, C.-W. J. Liu, M. Szurek, and N. Fakhri, Measuring irreversibility from learned representations of biological patterns, PRX Life **2**, 033013 (2024).
[48] I. Loshchilov and F. Hutter, Decoupled weight decay regularization, arXiv:1711.05101.
[49] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.
[50] Y. Wang and C.-Y. Lai, Multi-stage neural networks: Function approximator of machine precision, J. Comput. Phys. **504**, 112865 (2024).
[51] J. Duchi, E. Hazan, and Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, J. Mach. Learn. Res. **12**, 2121 (2011).
[52] J. Page, Super-resolution of turbulence with dynamics in the loss, J. Fluid Mech. **1002**, R3 (2025).
[53] R. Durall, M. Keuper, and J. Keuper, Watch your up-convolution: CNN based generative deep neural networks are failing to reproduce spectral distributions, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 7887 (IEEE, New York, 2020).
[54] J. Ng, Y. Wang, and C.-Y. Lai, Spectrum-informed multistage neural networks: Multiscale function approximators of machine precision, arXiv:2407.17213.
[55] K. Pearson, On lines and planes of closest fit to systems of points in space, London, Edinburgh, Dublin Philos. Mag. J. Sci. **2**, 559 (1901).
[56] A. Vela-Martín, Complexity of extreme-event prediction in turbulent flows, Phys. Rev. Fluids **9**, 104603 (2024).
[57] D. Lucas and R. R. Kerswell, Recurrent flow analysis in spatiotemporally chaotic 2-dimensional Kolmogorov flow, Phys. Fluids **27**, 045106 (2015).
[58] L. van der Maaten and G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. **9**, 2579 (2008).
[59] D. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley Series in Probability and Statistics (Wiley, New York, 2015).
[60] P. Cvitanović and J. F. Gibson, Geometry of the turbulence in wall-bounded shear flows: Periodic orbits, Phys. Scr. **T142**, 014007 (2010).
[61] G. Kawahara, M. Uhlmann, and L. van Veen, The significance of simple invariant solutions in turbulent flows, Annu. Rev. Fluid Mech. **44**, 203 (2012).
[62] M. D. Graham and D. Floryan, Exact coherent states and the nonlinear dynamics of wall-bounded turbulent flows, Annu. Rev. Fluid Mech. **53**, 227 (2021).
[63] R. Artuso, E. Aurell, and P. Cvitanovic, Recycling of strange sets: I. Cycle expansions, Nonlinearity **3**, 325 (1990).
[64] R. Artuso, E. Aurell, and P. Cvitanovic, Recycling of strange sets: II. Applications, Nonlinearity **3**, 361 (1990).
[65] M. C. Krygier, J. L. Pughe-Sanford, and R. O. Grigoriev, Exact coherent structures and shadowing in turbulent Taylor–Couette flow, J. Fluid Mech. **923**, A7 (2021).
[66] C. C. Lalescu, C. Meneveau, and G. L. Eyink, Synchronization of chaos in fully developed turbulence, Phys. Rev. Lett. **110**, 084102 (2013).
[67] T. A. Zaki, Turbulence from an observer perspective, Annu. Rev. Fluid Mech. **57**, 311 (2024).
[68] http://www.cirrus.ac.uk.
[69] D. Hendrycks and K. Gimpel, Gaussian error linear units (GELUs), arXiv:1606.08415.
[70] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in *Proceedings of the 32nd International Conference on Machine Learning*, edited by F. Bach and D. Blei, Proceedings of Machine Learning Research Vol. 37 (PMLR, Lille, France, 2015), pp. 448–456.