Spectrally Regularized Latent Flow Matching for Turbulence Generation

Khalid Rafiq [1] **Aditya G. Nair** [1]

Abstract

Latent diffusion and flow matching have emerged as leading approaches for synthetic turbulence generation, yet they systematically under-represent dissipation-range amplitudes. We introduce a latent flow matching framework with a spectrally regularized compression stage that directly targets this failure mode. On a 256^2 DNS dataset at $Re_f \approx 2250$, replacing an MSE-trained VAE with a zone-weighted log-spectral objective raises deep-dissipation retained spectral power from 25% to 94% in reconstruction and from 20% to 79% in unconditional generation. The improved latent representation also yields a substantially better sampling cost–fidelity tradeoff: the MSE-trained latent space imposes a fundamental quality ceiling near DD bias -0.70 that no integrator or step-count can overcome, while the spectrally regularized latent space reaches DD bias -0.117 at just 20 function evaluations. Mechanistically, encoder–decoder swap experiments show that the improvement is driven primarily by encoder-induced latent reorganization rather than decoder capacity, while a support-amplitude decomposition reveals that MSE-trained models behave as conservative suppression models, minimizing pointwise error by attenuating intermittent high-wavenumber structure. Both pipelines recover the second-order structure function and the correct sign of S_3, indicating the correct cascade direction without explicit supervision. A small residual gap in the magnitude of S_3 suggests that phase-coherent triadic organization remains a complementary axis to amplitude fidelity for future generative turbulence models.

[1]Department of Mechanical Engineering, University of Nevada, Reno, NV, USA. Correspondence to: Khalid Rafiq <krafiq@unr.edu>.

Proceedings of the AI4Physics Workshop at the 43rd International Conference on Machine Learning (AI4Physics@ICML 2026), Seoul, South Korea. 2026. Copyright 2026 by the author(s).

1. Introduction

A practical generator of turbulent flow fields would substantially reduce the cost of downstream tasks that currently rely on direct numerical simulation (DNS), including uncertainty quantification, ensemble statistics, closure-model training, and synthetic inflow generation. Recent advances have converged on latent generative pipelines, where diffusion or flow matching models are trained on learned low-dimensional representations of turbulent states (Rombach et al., 2022; Lipman et al., 2023; Du et al., 2024; Liu et al., 2025; Parikh et al., 2026; Oommen et al., 2024; Whittaker et al., 2024; Granero Belinchon & Cabeza Gallucci, 2024). Despite their success, these models exhibit a persistent failure mode: when trained with pointwise reconstruction objectives, they systematically under-represent dissipation-range amplitudes. This limitation is particularly important because high-wavenumber dynamics govern enstrophy dissipation and strongly influence the evolution of downstream flow physics (Davidson, 2015). More broadly, the role of the compression objective in shaping downstream generative dynamics remains poorly understood. In latent generative models, the encoder does not merely compress the data distribution; it organizes the geometry of the latent manifold on which sampling and transport are subsequently performed. We show that modifying this compression objective can substantially alter both generative fidelity and sampling efficiency.

We introduce a latent flow matching framework with a spectrally regularized compression stage based on a zone-weighted log-spectral objective. The learned latent representation is propagated through a CondOT-path flow matching generator (Lipman et al., 2023). On forced two-dimensional Navier–Stokes turbulence at $Re_f \approx 2250$ on a 256^2 grid, replacing an MSE-trained variational autoencoder with the proposed spectral objective improves deep-dissipation retained spectral power from 25% to 94% in reconstruction and from 20% to 79% in unconditional generation. The improvement also reshapes the latent transport geometry: the MSE-trained latent space imposes a fundamental quality ceiling near DD bias -0.70 that no integrator or step-count can overcome, while the spectrally regularized latent space reaches DD bias -0.117 at just 20 function evaluations.

To understand the source of this gain, we introduce two diag-

nostics that probe how compression objectives reshape latent representations. First, an encoder–decoder swap experiment shows that the improvement is co-adapted, with encoder-side latent reorganization playing the dominant role; the baseline decoder is unable to interpret the reorganized latent code. Second, a support–amplitude decomposition reveals a structural failure mode of pointwise reconstruction losses: MSE-trained models act as conservative suppression models, achieving low pointwise error by systematically attenuating intermittent dissipation-range structures. In contrast, the spectral objective restores high-wavenumber amplitudes at a small pointwise cost. Together, these diagnostics identify where spectral regularization acts within the architecture and clarify how reconstruction objectives influence downstream generative behavior.

The same controlled setup also clarifies what the method does not resolve. Both pipelines recover the second-order structure function $S_2(r)$ and the correct *sign* of the third-order structure function $S_3(r)$, and hence the direction of the cascade, without explicit supervision on structure statistics. However, the *magnitude* of $S_3(r)$ retains a small residual gap that the spectral regularizer does not close. This is consistent with the construction of the objective itself: shell-averaged spectral penalties constrain Fourier amplitudes but are intrinsically insensitive to inter-scale phase organization and triadic coherence. We therefore view phase-coherent interactions as a complementary axis to amplitude fidelity rather than a competing one.

Contributions. We design a controlled two-pipeline study in which all architectural components are held fixed except the compression objective, enabling direct isolation of spectral regularization effects. This yields the following contributions:

1. **Spectrally consistent generative modeling.** Spectral regularization substantially improves recovery of fine-scale structure in unconditional latent generation, raising deep-dissipation retained spectral power from 25% to 94% in reconstruction and from 20% to 79% in generation.

2. **Improved sampling efficiency through latent geometry.** The MSE-trained latent space imposes a fundamental quality ceiling that no integrator or step-count can overcome; the spectrally regularized latent manifold accesses a strictly better fidelity regime at as few as 20 function evaluations.

3. **Mechanistic understanding of latent reorganization.** Encoder–decoder swap experiments demonstrate that the gains arise primarily from encoder-driven latent reorganization rather than increased decoder expressivity.

4. **A failure mode of pointwise reconstruction losses.** Pointwise objectives favor suppression of intermittent high-frequency structure, leading to deceptively low reconstruction error despite poor spectral fidelity.

5. **Phase coherence as a complementary generative axis.** Both pipelines recover S_2 and the sign of S_3 without explicit supervision, while the residual gap in the magnitude of S_3 identifies phase-coherent triadic organization as a natural direction for future generative objectives.

2. Related Work

Latent generative models for turbulence. Deep generative models have been increasingly applied to turbulent flow fields. Early work focused on GANs (Drygala et al., 2022; Granero Belinchon & Cabeza Gallucci, 2024), while more recent approaches are dominated by diffusion (Ho et al., 2020; Song et al., 2021) and flow-matching models (Lipman et al., 2023; Liu et al., 2023; Albergo & Vanden-Eijnden, 2023; Shaul et al., 2023). CoNFiLD (Du et al., 2024) and CoNFiLD-inlet (Liu et al., 2025) combine latent diffusion with neural-field representations of fluid flows, while Parikh et al. (2026) apply conditional flow matching to wall-bounded turbulence. Whittaker et al. (2024) further show that diffusion models can recover inertial-range scaling statistics. Hybrid approaches combining neural operators with diffusion have also been proposed to improve spectral accuracy (Oommen et al., 2024).

Across these works, latent compression plays a central role (Nakamura et al., 2021; 2020; Doan et al., 2021; 2023; Wiewel et al., 2019; Rafiq & Nair, 2025; Rafiq et al., 2025; Kontolati et al., 2023). However, existing pipelines rely on pointwise (MSE-based) reconstruction objectives, which systematically under-resolve the dissipation range of the spectrum. The present work addresses this limitation by introducing a spectrally regularized compression stage within a latent flow matching pipeline and showing that the resulting gain propagates from reconstruction to unconditional generation.

Spectral and physics-aware losses. Neural networks exhibit a well-documented bias toward low-frequency content (Rahaman et al., 2019), motivating the use of frequency-aware objectives. Spectral losses have been explored in neural-operator forecasting of chaotic systems (Chakraborty et al., 2026) and in diffusion models augmented with operator structure (Oommen et al., 2024).

Our approach differs in two key ways. First, spectral regularization is applied at the compression bottleneck of a generative model, rather than as a forecasting penalty. Second, the loss is zone-weighted in log-spectral space to explicitly target the disparity between inertial-range and dissipation-range amplitudes. The encoder–decoder swap and support–

amplitude diagnostics introduced here further provide a mechanistic analysis of how reconstruction objectives reshape latent geometry and downstream generative behavior, which has not been examined in prior work.

Subgrid-scale modeling and turbulence super-resolution. Two adjacent threads tackle related problems of recovering unresolved high-wavenumber content. Data-driven subgrid-scale (SGS) closure for large-eddy simulation seeks to express the effect of unresolved scales on resolved ones via neural surrogates. Representative approaches train neural surrogates on filtered DNS to recover unresolved stresses (Maulik & San, 2017; Gamahara & Hattori, 2017; Beck et al., 2019; Sirignano et al., 2020); later *a-posteriori*-stable formulations target consistency with the LES rollout itself, including stability-constrained and physics-consistent parameterizations for 2D and quasi-geostrophic turbulence (Park & Choi, 2021; Guan et al., 2022; Frezat et al., 2022; Kurz et al., 2023). Turbulence super-resolution casts scale recovery as supervised upsampling from a coarse field. Multi-scale CNN super-resolvers (Fukami et al., 2019) have been extended with physics-informed and adversarial losses (Bode et al., 2021; Kim et al., 2021; Yousif et al., 2023), and recent diffusion- and score-based methods have been applied to turbulent super-resolution and reconstruction (Shu et al., 2023; Oommen et al., 2024).

Our setting differs from both threads in two key ways. First, we study *unconditional* generation rather than coarse-to-fine reconstruction: dissipation-range structure must emerge from the latent prior alone. Second, the failure mode we address is structural rather than algorithmic. The high-k attenuation observed in latent generators is induced by the pointwise reconstruction objective at the compression bottleneck itself, not by filtered-DNS training or the choice of downsampling operator. The diagnostics introduced here (spectral-bias zoning, encoder–decoder swap, and support–amplitude decomposition) generalize naturally to both settings.

2D turbulence and phase organization. The classical theory of two-dimensional turbulence (Kraichnan, 1967; Batchelor, 1969; Kolmogorov, 1941; Frisch, 1995; Boffetta & Ecke, 2012; Davidson, 2015) describes an inverse energy cascade and a direct enstrophy cascade mediated by vortex filamentation. This process is intrinsically phase-coherent: strain induced by one vortex deforms another into elongated structures. The third-order longitudinal structure function $S_3(r)$ provides a standard diagnostic of this inter-scale organization.

Causality and controllable structures. Prior work has identified the structures that dominate interaction and control in 2D turbulence. Jiménez (2018) and Jiménez (2020) show via perturbation experiments that vortex dipoles, rather than individual vortices, are the most causally significant structures under velocity-based norms. Similarly, Yeh et al. (2021) use network-based analysis of Navier–Stokes interactions to identify dipole regions as dominant broadcast sites for perturbations and control. These results emphasize that generators intended for downstream control or causality use must preserve the joint amplitude–phase organization of vortex pairs, motivating our use of S_2 and S_3 as complementary diagnostics alongside the spectrum.

3. Problem Setup and Data

We solve the 2D incompressible Navier–Stokes equations in vorticity form on the doubly periodic domain $\Omega = [0, 2\pi]^2$,

$$\partial_t \omega + (\mathbf{u}\cdot\nabla)\omega = \nu\nabla^2\omega + f, \quad \nabla\cdot\mathbf{u} = 0, \tag{1}$$

with $\nu = 10^{-3}$ and a body force injecting energy near $k_f = 4$. Simulations are performed using `jax-cfd` (Kochkov et al., 2021) on a 256^2 grid with fourth-order time integration and standard $2/3$ dealiasing, yielding $k_{\max} = 85$. The dataset is integrated for $T = 600$ ($1.2{\times}10^6$ steps), with snapshots saved every 200 steps. The first 1000 snapshots are discarded as transient spin-up, after which the flow reaches a statistically stationary regime. The remaining 5000 fields are split temporally into 4500 train and 500 test, and standardized using training statistics.

The forcing-scale Reynolds number is $Re_f = u_{\mathrm{rms}} L_f/\nu \approx 2.25\times10^3$ with $u_{\mathrm{rms}} \approx 1.43$, placing the run in a moderate-Reynolds turbulent regime. The forcing-scale turnover time is $\tau_f = L_f/u_{\mathrm{rms}} \approx 1.10$, so the retained dataset spans approximately $456\,\tau_f$. The moderate scale separation $k_{\max}/k_f \approx 21$ produces a resolved but finite cascade range, consistent with prior studies of moderate-Reynolds-number two-dimensional turbulence (Boffetta & Ecke, 2012).

Spectral zoning. We partition the resolved spectrum into three bands used throughout the paper, illustrated in Figure 1. The inertial-range band IR ($k = 6$–40) contains the dominant large-scale coherent structures visible in the green-filtered inset of Figure 1, while the dissipation-onset band DO ($k = 41$–65) and the deep-dissipation band DD ($k = 66$–85) progressively isolate finer-scale intermittent structures. This spectral separation induces a severe imbalance under pointwise reconstruction objectives. Characteristic IR vorticity amplitudes are $O(\pm 7.5)$, whereas DD amplitudes are only $O(\pm 0.4)$, corresponding to an approximate $20\times$ disparity in signal magnitude, as shown by the filtered realizations in Figure 1. Under an ℓ_2 reconstruction loss, this becomes an effective $\sim 400\times$ imbalance in squared-error weighting per spatial location, causing standard objectives to overwhelmingly prioritize large-scale structure while systematically underweighting fine-scale

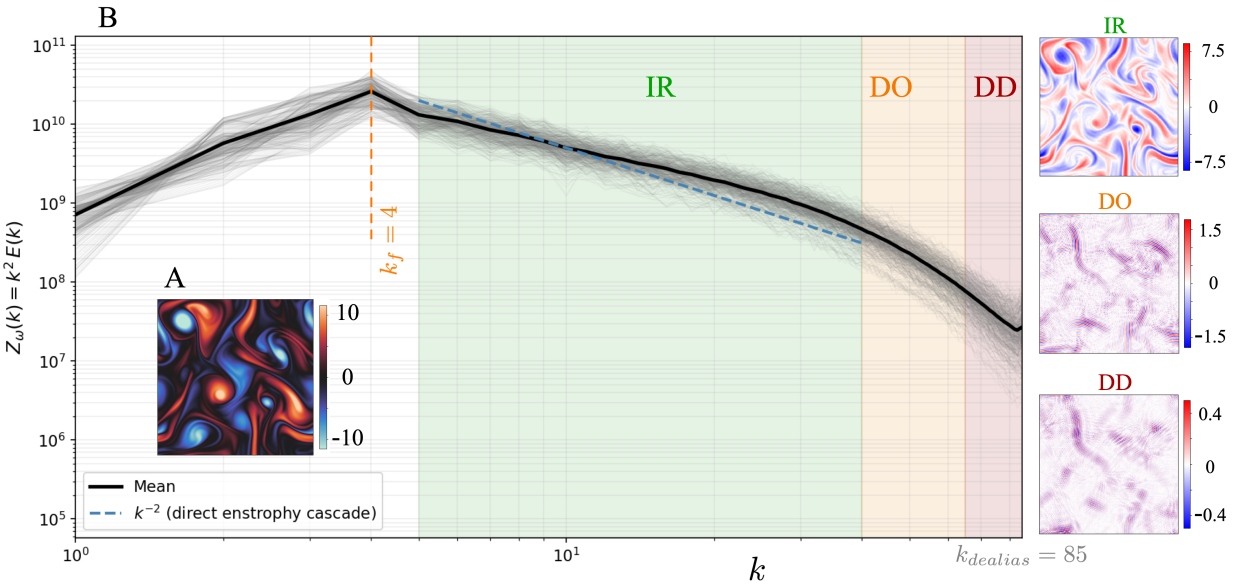

Figure 1. Dataset overview and spectral zoning. **(A)** Representative standardized vorticity snapshot from the statistically stationary 256^2 DNS at $Re_f \approx 2250$. **(B)** Mean shell-averaged vorticity power $Z_\omega(k) = k^2 E(k)$ over the test set (black) with individual realization spectra (grey) and the k^{-2} direct-enstrophy-cascade reference slope (blue dashed). The forcing wavenumber $k_f = 4$ and dealiasing cutoff $k_{\text{dealias}} = 85$ are marked. Shaded regions indicate the three spectral zones used throughout the paper: an inertial-range band IR ($k = 6$–40, green), a dissipation-onset band DO ($k = 41$–65, orange), and a deep-dissipation band DD ($k = 66$–85, red). Right panels show representative band-pass-filtered realizations within each spectral zone. The characteristic amplitude decreases from $O(\pm 7.5)$ in IR to $O(\pm 0.4)$ in DD, producing an approximate $20\times$ signal imbalance across scales and an effective $\sim 400\times$ disparity under pointwise ℓ_2 reconstruction losses.

content. This imbalance motivates the zone-weighted spectral regularization introduced in this work.

4. Method

4.1. Two-stage Pipeline

The framework, summarized in Figure 2, separates representation learning from latent generative transport. Stage 1 is a residual VAE that maps a vorticity snapshot $\omega \in \mathbb{R}^{1\times 256\times 256}$ to a structured latent tensor $z \in \mathbb{R}^{8\times 16\times 16}$ (a $32\times$ compression in spatial volume) and reconstructs back. Stage 2 freezes the decoder, encodes the training set with the encoder mean $\mu_\phi(\omega)$ (Rombach et al., 2022), and learns an unconditional flow-matching model on the resulting latent representation manifold. Sampling is performed by integrating the learned vector field forward from a Gaussian prior and decoding the terminal latent state through the frozen Stage 1 decoder. The key design hypothesis is that modifying the compression objective reshapes the geometry of the latent representation on which the flow matching dynamics are subsequently learned.

We instantiate two compression models with identical architecture and hyperparameters and differing only in objective. Both encoder and decoder are multiresolution residual convolutional networks (He et al., 2016; Wu & He, 2018; Hendrycks & Gimpel, 2016) with circular padding to respect periodicity. Both have $\sim$5M parameters. The Stage 2 velocity field $v_\theta(z, \tau)$ is a lightweight U-Net (Ronneberger et al., 2015) of $\sim$2M parameters with sinusoidal time embedding and self-attention at the bottleneck. Architecture details appear in Section A.

4.2. Compression Objectives

Model A (baseline). Model A uses the standard VAE objective:

$$\mathcal{L}_A = \frac{1}{N}\sum_j \left[\, \|\omega_j - \hat{\omega}_j\|_2^2 + \beta\, \mathrm{KL}\big(q_\phi(z|\omega_j)\,\|\,\mathcal{N}(0, I)\big)\right]. \tag{2}$$

Model B (zone-weighted log-spectral). Model B augments Equation (2) with a shell-wise spectral penalty. For each integer wavenumber shell $\mathcal{S}_k$ define the shell-averaged vorticity power

$$Z_\omega(k) = \frac{1}{|\mathcal{S}_k|}\sum_{(k_x,k_y)\in\mathcal{S}_k} |\hat{\omega}(k_x, k_y)|^2. \tag{3}$$

For zone $\mathcal{K}_z \in \{\mathrm{IR}, \mathrm{DO}, \mathrm{DD}\}$ the zone loss is the mean squared log-spectral error,

$$\mathcal{L}_z = \frac{1}{|\mathcal{K}_z|}\sum_{k\in\mathcal{K}_z} \left[\log(Z_{\hat{\omega}}(k)+\epsilon) - \log(Z_\omega(k)+\epsilon)\right]^2, \tag{4}$$

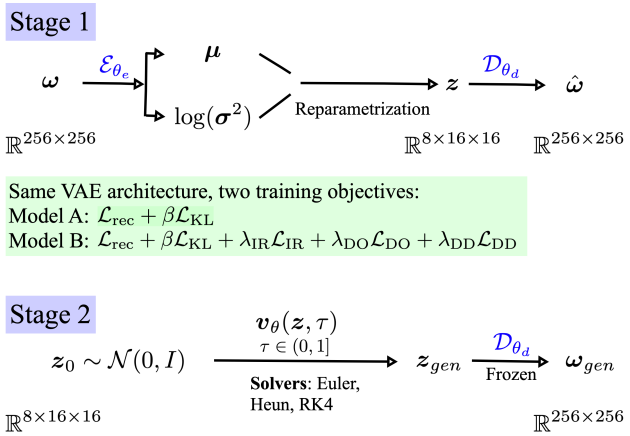

Figure 2. Two-stage latent generative pipeline. Stage 1 is a residual VAE trained with either a baseline objective (Model A) or a zone-weighted spectral objective (Model B). Stage 2 freezes the decoder and trains a latent CondOT flow matching generator. Sampled latents are decoded back to vorticity. The two pipelines share architecture, dataset, and Stage 2 generator exactly, differing only in compression objective and the resulting latent geometry.

giving the full objective

$$\mathcal{L}_B = \mathcal{L}_A + \lambda_{\text{IR}}\mathcal{L}_{\text{IR}} + \lambda_{\text{DO}}\mathcal{L}_{\text{DO}} + \lambda_{\text{DD}}\mathcal{L}_{\text{DD}}. \quad (5)$$

Equation (4) constrains the modulus of the Fourier transform on shells of constant $|\mathbf{k}|$. It is invariant to (a) the phase of any individual mode, (b) the relative phase between modes, and (c) the distribution of energy among modes within a shell. It cannot enforce inter-scale phase organization or triadic coherence.

4.3. Latent Flow Matching

We use the linear conditional optimal transport (CondOT) probability path (Lipman et al., 2023). Given $z_1 \sim q(z)$, $\varepsilon \sim \mathcal{N}(0, I)$, $\tau \sim \mathcal{U}[0, 1]$, the interpolant $z_\tau = (1-\tau)\varepsilon + \tau z_1$ has target velocity $u_\tau(z_\tau|z_1) = z_1 - \varepsilon$, and the loss is

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{\tau, z_1, \varepsilon}\left\|v_\theta(z_\tau, \tau) - (z_1 - \varepsilon)\right\|_2^2. \quad (6)$$

At inference, $z_0 \sim \mathcal{N}(0, T^2 I)$ is integrated forward by

$$\frac{dz_\tau}{d\tau} = v_\theta(z_\tau, \tau), \quad \tau \in [0, 1], \quad (7)$$

and the terminal state is decoded. The scalar T corrects a mild empirical under-dispersion fitted from 500 draws from the Gaussian prior $\mathcal{N}(0, I)$. The nearly identical values $T_A = 1.157$ and $T_B = 1.170$ confirm the calibration does not explain the gap between pipelines. CondOT is chosen because its low-kinetic-energy trajectories enable accurate generation at low NFE (Shaul et al., 2023).

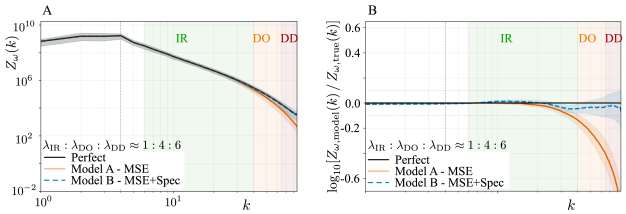

Figure 3. Stage 1 ensemble spectral fidelity. **Left:** mean shell-averaged $Z_\omega(k)$ over the held-out test set with 10–90% bands. **Right:** spectral bias mean $\pm 1\sigma$. Spectral regularization in Model B moves the bias closer to zero in all three zones, with the largest gain in DD. Both models recover the IR slope.

Table 1. Zone-averaged spectral statistics for the two compression models on the held-out test set. "ret." is the retained spectral power $100 \times 10^{\text{bias}}$ (closer to 100% is better).

	IR			DO			DD		
	bias	spr.	ret.	bias	spr.	ret.	bias	spr.	ret.
A	−0.04	0.05	90.8	−0.27	0.11	54.1	−0.61	0.19	24.8
B	**−0.01**	**0.03**	**97.1**	**−0.03**	**0.06**	**92.3**	**−0.03**	**0.11**	**93.6**

Training. Both VAEs are trained for 200 epochs with batch size 48, learning rate 7.5×10^{-4}, and $\beta = 7.5 \times 10^{-3}$. For Model B, spectral weights are selected via Bayesian sweep, yielding the ratio $1 : 4 : 6$ for $(\lambda_{\text{IR}}, \lambda_{\text{DO}}, \lambda_{\text{DD}})$. A uniform spectral weighting ($\lambda_{\text{IR}} = \lambda_{\text{DO}} = \lambda_{\text{DD}} = 4 \times 10^{-3}$) was also evaluated and improved over the MSE-only baseline, but the zone-weighted $(1:4:6)$ configuration achieves lower bias across all three spectral zones. Stage 2 flow-matching models are trained for 1000 epochs with batch size 64 and learning rate 3×10^{-4}.

5. Stage 1: Effects of Spectral Regularization

5.1. Ensemble Spectral Fidelity

Figure 3 reports the held-out shell-averaged vorticity spectrum and per-wavenumber bias $\log_{10}[Z_{\omega,\text{model}}(k)/Z_{\omega,\text{true}}(k)]$. Model A exhibits a small negative bias in IR and a strong high-k deficit; Model B is close to zero across all three zones with narrower uncertainty. Quantitatively (Table 1), Model A preserves only 54% of true DO and 25% of true DD power, while Model B preserves 92% and 93% respectively.

5.2. Suppression vs. Recovery in the Dissipation Range

Better Fourier-space fidelity does not imply lower pointwise error in the DD band. Across the test ensemble Model B has a slightly but systematically *larger* DD-band MSE than Model A ($\mu = 6.7 \times 10^{-3}$ vs. 6.2×10^{-3}); see Section B. To explain this dissociation, we threshold the band-pass DD field at the p-th percentile of true DD magnitude, producing binary support masks, and decompose model predictions

into true positives, false negatives, and false positives on the union of active support (Figure 4).

The two pipelines behave qualitatively differently. Model A acts as a *conservative-suppression* model: predicting near-zero in sparse DD regions minimizes MSE with little penalty, systematically suppressing true support and amplitudes by $\sim 2\times$. Model B behaves as a *recovery* model, restoring most of the true support and amplitude budget at the cost of a slightly larger pointwise error. Low MSE on intermittent fine-scale structure can therefore reflect suppression rather than faithful recovery.

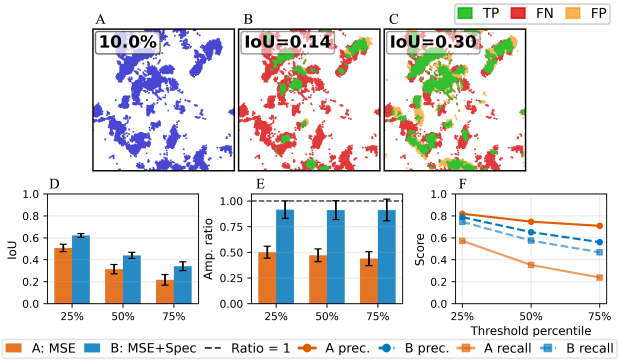

Figure 4. Support–amplitude decomposition in the DD band ($k = 66$–85). **(A)** Thresholded DD-support mask from one true field. **(B–C)** Predictions of Models A and B as TP (green), FN (red), FP (orange). **(D)** Ensemble IoU across thresholds. **(E)** Amplitude ratio (model/truth) on the union support. **(F)** Precision and recall. Model A is precise but strongly suppressed (amplitude ratio ≈ 0.44); Model B has higher IoU, much higher recall, and an amplitude ratio ≈ 0.91 at the cost of broader support.

5.3. Where Does the Spectral Gain Live?

To localize the spectral benefit, we evaluate all four pairings of $\{\mathcal{E}_A, \mathcal{E}_B\} \times \{\mathcal{D}_A, \mathcal{D}_B\}$ (Figure 5, Table 2). The matched spectrally regularized pair $\mathcal{D}_B \circ \mathcal{E}_B$ is the only configuration that holds low bias across all three zones. Cross-swapped pairs are catastrophic or partial: $\mathcal{D}_A \circ \mathcal{E}_B$ is worse than the baseline in every zone (DD bias -0.96), indicating that $\mathcal{E}_B$'s latent representation is reorganized into a form $\mathcal{D}_A$ cannot decode. The reverse swap $\mathcal{D}_B \circ \mathcal{E}_A$ partially recovers DD (-0.23 vs. -0.61) but is worse than baseline in IR and DO. The spectral benefit is therefore co-adapted, with an asymmetric encoder-side anchor: encoder-side latent reorganization is the more fundamental component, and the decoder contributes complementary but limited recovery capacity.

6. Stage 2: Spectrally Faithful Generation at Low Cost

6.1. Generated Spectra Inherit the Stage 1 Gain

Figure 6 (top) shows representative unconditional samples from both pipelines together with ensemble shell-averaged spectra over 500 generated and 500 test fields. Both pipelines produce visually plausible fields with realistic large-scale organization. The spectra echo the Stage 1 result: Model A holds the dominant scales but develops a strong high-k deficit; Model B remains close to truth across all three zones (IR bias -0.03 vs. -0.10, DO -0.10 vs. -0.36, DD -0.10 vs. -0.70). The spectral regularization of the encoder propagates through to the generative distribution, raising DD retained spectral power from 20% to 79% in generation.

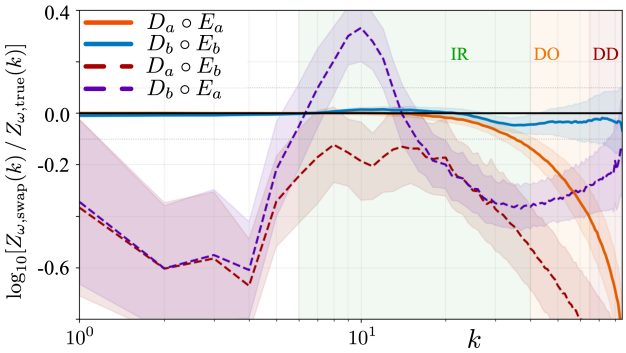

Figure 5. Encoder–decoder swap diagnostic. Spectral bias for the four combinations of $\{\mathcal{E}_A, \mathcal{E}_B\} \times \{\mathcal{D}_A, \mathcal{D}_B\}$. Only the matched pair $\mathcal{D}_B \circ \mathcal{E}_B$ holds low bias across all zones. The mismatch $\mathcal{D}_A \circ \mathcal{E}_B$ catastrophically degrades, showing that the baseline decoder cannot reliably interpret the spectrally regularized latent representation.

Table 2. Zone-averaged spectral bias for all four encoder–decoder pairings.

	IR		DO		DD	
	bias	spr.	bias	spr.	bias	spr.
$\mathcal{D}_A \circ \mathcal{E}_A$	-0.042	0.049	-0.267	0.114	-0.606	0.192
$\mathcal{D}_B \circ \mathcal{E}_B$	**-0.013**	**0.031**	**-0.035**	**0.062**	**-0.029**	**0.112**
$\mathcal{D}_A \circ \mathcal{E}_B$	-0.286	0.198	-0.702	0.203	-0.961	0.196
$\mathcal{D}_B \circ \mathcal{E}_A$	-0.171	0.243	-0.321	0.098	-0.228	0.150

6.2. Sampling Cost–Fidelity and Latent Quality Ceiling

Figure 7 reports the Stage 2 spectral bias as a function of the number of function evaluations (NFE) for fixed-step Euler and Heun integrators, by zone. On an NVIDIA RTX 3090 with batch 50, generation costs ~ 3.4 ms per NFE. Across all NFE budgets and all three zones, Model B has a strictly more favourable cost–fidelity curve than Model A. Critically, Model A Heun saturates near DD bias -0.70 from NFE=20 onwards, matching Model A Euler at convergence; indicating that the MSE-trained latent space imposes a fundamental quality ceiling that no integrator or step-count can overcome. Model B Heun reaches DD bias -0.117 at NFE=20 and holds it, accessing a fidelity regime entirely unavailable to Model A. Heun reaches this converged accuracy at a small fraction of the Euler cost; RK4 produces

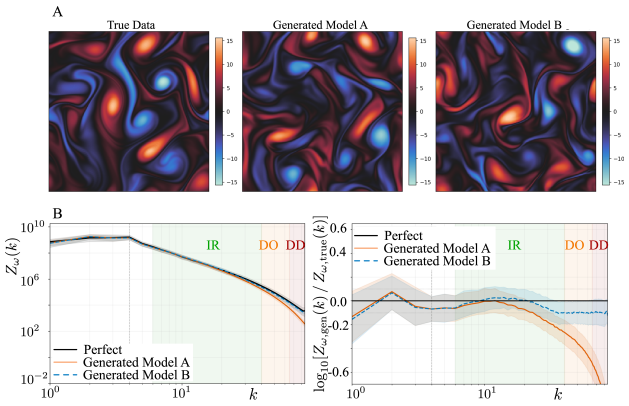

Figure 6. Stage 2 generation. **Top:** representative held-out test field and unconditional samples from each pipeline. The samples are draws from the learned latent distribution, not reconstructions. **Bottom:** mean shell-averaged spectrum and bias over 500 generated samples vs. 500 test fields.

Table 3. Stage 2 generative spectral bias and retained power over 500 generated vs. 500 test fields.

	IR			DO			DD		
	bias	spr.	ret.	bias	spr.	ret.	bias	spr.	ret.
A	−0.10	0.13	79.8	−0.36	0.14	43.8	−0.70	0.20	20.0
B	**−0.03**	**0.10**	**92.5**	**−0.10**	**0.10**	**79.6**	**−0.10**	**0.11**	**79.4**

results indistinguishable from Heun at equal NFE and is omitted, consistent with the smooth low-kinetic trajectories of the CondOT path (Shaul et al., 2023). Model B with Heun at NFE=20 is the practical sweet spot.

The result is particularly striking because the Stage 2 generators are architecturally identical; only the loss function changes which affects the latent geometry. The ceiling reflects a representation bottleneck rather than an optimization failure: Model A's Stage 1 reconstructions exhibit the same DD deficit (Table 1), confirming that fine-scale information is absent from the latent code and cannot be recovered by any downstream sampler.

6.3. Structure Functions: Amplitude and Phase

The same controlled setup that delivers the amplitude and cost–fidelity gains lets us probe the phase content of the generated fields, which the shell-averaged regularizer does not constrain by construction. We compute longitudinal velocity-increment structure functions $S_p(r) = \langle[\delta u_L(\mathbf{x}, r)]^p\rangle$ for $p = 2, 3$, recovering velocity from generated vorticity via $\hat{\psi} = -\hat{\omega}/k^2$ and averaging over 24 azimuthal directions and 500 realizations. Here $S_2(r)$ and $S_3(r)$ are the second- and third-order longitudinal velocity-increment structure functions; S_3 is negative in 2D turbulence and its sign encodes the direction of the enstrophy cascade. Two cautions apply. First, dividing by k^2 in

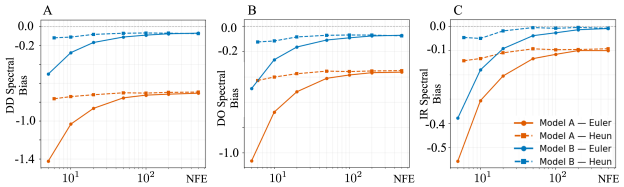

Figure 7. Sampling cost–fidelity tradeoff. Spectral bias vs. NFE in the DD (left), DO (middle), and IR (right) zones. Solid: Euler. Dashed: Heun. Model A saturates near DD bias -0.70 regardless of integrator or NFE, revealing a fundamental quality ceiling of the MSE-trained latent space. Model B Heun reaches DD bias -0.117 at NFE=20; a fidelity level entirely inaccessible to Model A at any sampling budget.

Fourier space damps high-k content in velocity, so velocity-increment statistics test a different aspect of the field than vorticity-shell spectra. Second, S_3 is sensitive to phase-coherent triadic interactions (Frisch, 1995; Davidson, 2015), the property a shell-averaged loss cannot enforce.

Figure 8 reports the result. Both pipelines reproduce $S_2(r)$ across the resolved range and recover the correct *sign* of $S_3(r)$ throughout the inertial range, i.e., the direction of the cascade, without any explicit supervision on structure functions. This indicates that the learned latent dynamics preserve sufficient multiscale organization for the correct cascade direction to emerge without explicit supervision. The magnitudes of S_2 and S_3 are also in close agreement with the test ensemble, with a small residual gap in the deepest part of S_3.

Reading the small gap. The construction of Equation (4) is informative here: a shell-averaged loss controls $|\hat{\omega}(\mathbf{k})|^2$ on annuli of constant $|\mathbf{k}|$ and is invariant to the relative phase between modes within a shell and across shells. We therefore expect the spectral regularizer to buy amplitude fidelity but not, by itself, to enforce inter-scale phase coherence. The fact that S_3 is nonetheless captured in sign and approximately in magnitude indicates that phase organization is largely recovered through the latent dynamics, while the residual gap is consistent with phase-coherent triadic interactions being only implicitly recovered through the MSE component of the objective (Chakraborty et al., 2026). Closing this remaining gap likely requires objectives that explicitly constrain phase-coherent multiscale interactions.

7. Discussion

Latent geometry and transport conditioning. The encoder–decoder swap experiments point to a broader observation about latent generative models for physical systems. The encoder does not merely compress the data distribution; it organizes the transport geometry on which the downstream generative dynamics are learned. Our results show

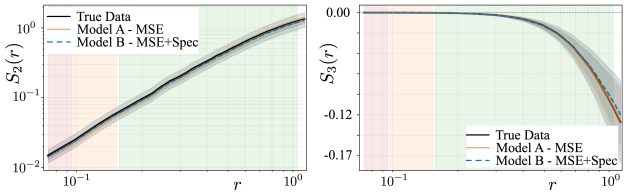

Figure 8. Longitudinal structure functions over 500 realizations. **Left:** $S_2(r)$. Both pipelines reproduce the second-order structure function across the resolved range. **Right:** $S_3(r)$. Both pipelines correctly recover the negative sign of S_3 and therefore the direction of the cascade without any structure-function supervision, with magnitudes close to the test ensemble. The shell-averaged regularizer is not designed to enforce phase coherence, so the small residual gap is an expected complementary direction rather than a failure of the present method.

that modifying the reconstruction objective reorganizes this latent geometry into a form that is substantially better conditioned for ODE-based transport and low-NFE sampling. The co-adaptation observed in the swap experiments further suggests that phase-sensitive refinement is most naturally introduced as a joint representation-learning objective, rather than as a correction applied only to the decoder or generative dynamics. Higher-order structure-function penalties, two-point correlation objectives, or bispectral constraints (Frisch, 1995) are natural extensions in this direction. More broadly, the results suggest that reconstruction objectives can substantially alter the conditioning of latent transport problems in generative models of multiscale physical systems. In the present setting, spectral regularization improves not only reconstruction fidelity but also the geometry on which latent transport trajectories are learned, leading to substantially improved low-NFE sampling behavior despite identical Stage 2 architectures.

Implications for downstream physical modeling. Generated turbulent ensembles are increasingly used for uncertainty quantification, surrogate training, synthetic inflow generation, flow control, and causality analysis. These downstream tasks depend not only on spectral amplitudes but also on the joint phase organization of coherent structures such as vortex dipoles. Our results show that spectrally regularized latent generation substantially improves fine-scale amplitude fidelity while preserving the correct cascade direction and large-scale structure-function behavior. The remaining gap in the magnitude of S_3 suggests that accurately reproducing phase-coherent multiscale interactions remains an important direction for future physics-aware generative objectives.

This distinction is particularly relevant for downstream analyses based on flow interactions rather than pointwise statistics. Prior work on broadcast-mode flow control (Yeh et al., 2021) and causality analysis (Jiménez, 2018; 2020) has shown that dynamically significant structures emerge from the collective organization of vortex pairs and interaction networks. Generative models intended for these applications must therefore recover not only spectral amplitudes but also the relative phase organization underlying coherent multiscale dynamics.

Limitations and future work. The present study uses a single Reynolds number ($Re_f \approx 2250$) on a 256^2 domain. Extending the analysis to higher Reynolds numbers, where the inertial range broadens and triadic interactions span larger spectral ranges, is a natural next step. The current implementation also uses a simple post-hoc variance calibration; more principled latent priors, latent whitening strategies, and adaptive transport paths remain promising directions for future work. The analysis is restricted to two-dimensional turbulence, where cascade structure and coherent interactions are comparatively well understood; extending to three-dimensional flows introduces vortex stretching and modified cascade dynamics, but does not alter the central observation: pointwise reconstruction objectives systematically bias the geometry of learned latent representations and underweight fine-scale multiscale structure; a bias that can be directly remedied by upweighting small-scale content in the spectral domain.

8. Conclusion

We introduced a spectrally regularized latent flow matching framework for generative modeling of multiscale physical fields and showed that modifying the compression objective substantially alters downstream generative behavior. A zone-weighted log-spectral regularizer at the latent bottleneck yields 93% retention of true DD spectral power in reconstruction and 79% in generation, while the MSE-trained latent space imposes a fundamental quality ceiling that no integrator or step-count can overcome.

Beyond improved spectral fidelity, the results reveal a mechanistic link between reconstruction objectives, latent geometry, and transport dynamics. Encoder–decoder swap experiments show that the gains arise primarily from encoder-driven latent reorganization, while the support–amplitude analysis demonstrates that pointwise reconstruction losses can achieve low error through systematic suppression of intermittent fine-scale structure. The spectrally regularized latent manifold is also substantially better conditioned for low-NFE ODE-based sampling despite identical Stage 2 architectures.

Both pipelines recover $S_2(r)$ and the correct sign of $S_3(r)$ without explicit supervision, while the remaining gap in S_3 magnitude highlights phase-coherent multiscale interactions as a natural target for future phase-aware objectives. More broadly, the results suggest that reconstruction objectives can fundamentally reshape latent transport geometry,

downstream sampling dynamics, and physical fidelity in generative models of multiscale systems.

Acknowledgements

A.G.N. acknowledges support from the National Science Foundation AI Institute in Dynamic Systems (Award No. 2112085, Program Manager: Dr. Shahab Shojaei-Zadeh). The authors gratefully acknowledge Professor Ping Liu from the Department of Computer Science and Engineering at the University of Nevada, Reno, for helpful and critical discussions.

Impact Statement

This work advances generative modeling methods for turbulent flow simulation. Improved fidelity in synthetic turbulence generation has potential benefits for downstream scientific applications including uncertainty quantification, climate modeling, and computational fluid dynamics, where high-fidelity simulation data is expensive to obtain. We do not foresee any negative societal impacts from this work.

References

Albergo, M. S. and Vanden-Eijnden, E. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations (ICLR)*, 2023. URL https://arxiv.org/abs/2209.15571.

Batchelor, G. K. Computation of the energy spectrum in homogeneous two-dimensional turbulence. *Physics of Fluids*, 12(Suppl. II):233–239, 1969.

Beck, A., Flad, D., and Munz, C.-D. Deep neural networks for data-driven LES closure models. *Journal of Computational Physics*, 398:108910, 2019. doi: 10.1016/j.jcp.2019.108910.

Bode, M., Gauding, M., Lian, Z., Denker, D., Davidovic, M., Kleinheinz, K., Jitsev, J., and Pitsch, H. Using physics-informed enhanced super-resolution generative adversarial networks for subfilter modeling in turbulent reactive flows. *Proceedings of the Combustion Institute*, 38(2): 2617–2625, 2021. doi: 10.1016/j.proci.2020.06.022.

Boffetta, G. and Ecke, R. E. Two-dimensional turbulence. *Annual Review of Fluid Mechanics*, 44:427–451, 2012. doi: 10.1146/annurev-fluid-120710-101240.

Chakraborty, D., Mohan, A. T., and Maulik, R. Binned spectral power loss for improved prediction of chaotic systems. *Journal of Computational Physics*, 558:114866, 2026. doi: 10.1016/j.jcp.2026.114866.

Davidson, P. A. *Turbulence: An Introduction for Scientists and Engineers*. Oxford University Press, 2nd edition, 2015.

Doan, N. A. K., Polifke, W., and Magri, L. Auto-encoded reservoir computing for turbulence learning. In *International Conference on Computational Science*, pp. 344–351. Springer, 2021.

Doan, N. A. K., Racca, A., and Magri, L. Convolutional autoencoder for the spatiotemporal latent representation of turbulence. In *International Conference on Computational Science*, pp. 328–335. Springer, 2023.

Drygala, C., Winhart, B., di Mare, F., and Gottschalk, H. Generative modeling of turbulence. *Physics of Fluids*, 34 (3), 2022.

Du, P., Parikh, M. H., Fan, X., Liu, X.-Y., and Wang, J.-X. Conditional neural field latent diffusion model for generating spatiotemporal turbulence. *Nature Communications*, 15(1):10416, 2024.

Frezat, H., Le Sommer, J., Fablet, R., Balarac, G., and Lguensat, R. A posteriori learning for quasi-geostrophic turbulence parametrization. *Journal of Advances in Modeling Earth Systems*, 14(11):e2022MS003124, 2022. doi: 10.1029/2022MS003124.

Frisch, U. *Turbulence: The Legacy of A.N. Kolmogorov*. Cambridge University Press, Cambridge, 1995. doi: 10.1017/CBO9781139170666.

Fukami, K., Fukagata, K., and Taira, K. Super-resolution reconstruction of turbulent flows with machine learning. *Journal of Fluid Mechanics*, 870:106–120, 2019. doi: 10.1017/jfm.2019.238.

Gamahara, M. and Hattori, Y. Searching for turbulence models by artificial neural network. *Physical Review Fluids*, 2(5):054604, 2017. doi: 10.1103/PhysRevFluids.2.054604.

Granero Belinchon, C. and Cabeza Gallucci, M. A multiscale and multicriteria generative adversarial network to synthesize 1-dimensional turbulent fields. *Machine Learning: Science and Technology*, 5(2):025032, 2024.

Guan, Y., Chattopadhyay, A., Subel, A., and Hassanzadeh, P. Stable a posteriori LES of 2D turbulence using convolutional neural networks: Backscattering analysis and generalization to higher Re via transfer learning. *Journal of Computational Physics*, 458:111090, 2022. doi: 10.1016/j.jcp.2022.111090.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *CVPR*, 2016.

Hendrycks, D. and Gimpel, K. Gaussian error linear units (GELUs). *arXiv preprint arXiv:1606.08415*, 2016.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 6840–6851, 2020. URL https://arxiv.org/abs/2006.11239.

Jiménez, J. Machine-aided turbulence theory. *Journal of Fluid Mechanics*, 854:R1, 2018. doi: 10.1017/jfm.2018.660.

Jiménez, J. Monte carlo science. *Journal of Turbulence*, 21(9–10):544–566, 2020. doi: 10.1080/14685248.2020.1813259.

Kim, H., Kim, J., Won, S., and Lee, C. Unsupervised deep learning for super-resolution reconstruction of turbulence. *Journal of Fluid Mechanics*, 910:A29, 2021. doi: 10.1017/jfm.2020.1028.

Kochkov, D., Smith, J. A., Alieva, A., Wang, Q., Brenner, M. P., and Hoyer, S. Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences*, 118(21):e2101784118, 2021. doi: 10.1073/pnas.2101784118.

Kolmogorov, A. N. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Doklady Akademii Nauk SSSR*, 30(4):301–305, 1941.

Kontolati, K., Goswami, S., Karniadakis, G. E., and Shields, M. D. Learning in latent spaces improves the predictive accuracy of deep neural operators. *arXiv preprint arXiv:2304.07599*, 2023.

Kraichnan, R. H. Inertial ranges in two-dimensional turbulence. *Physics of Fluids*, 10(7):1417–1423, 1967. doi: 10.1063/1.1762301.

Kurz, M., Offenhäuser, P., and Beck, A. Deep reinforcement learning for turbulence modeling in large eddy simulations. *International Journal of Heat and Fluid Flow*, 99:109094, 2023. doi: 10.1016/j.ijheatfluidflow.2022.109094.

Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations (ICLR)*, 2023. URL https://arxiv.org/abs/2210.02747.

Liu, X., Gong, C., and Liu, Q. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations (ICLR)*, 2023. URL https://arxiv.org/abs/2209.03003.

Liu, X.-Y., Parikh, M. H., Fan, X., Du, P., Wang, Q., Chen, Y.-F., and Wang, J.-X. Confild-inlet: Synthetic turbulence inflow using generative latent diffusion models with neural fields. *Physical Review Fluids*, 10(5):054901, 2025.

Maulik, R. and San, O. A neural network approach for the blind deconvolution of turbulent flows. *Journal of Fluid Mechanics*, 831:151–181, 2017. doi: 10.1017/jfm.2017.637.

Nakamura, T., Fukami, K., Hasegawa, K., Nabae, Y., and Fukagata, K. Cnn-ae/lstm based turbulent flow forecast on low-dimensional latent space. In *3rd Workshop on Machine Learning and the Physical Sciences at NeurIPS*, pp. p–8, 2020.

Nakamura, T., Fukami, K., Hasegawa, K., Nabae, Y., and Fukagata, K. Convolutional neural network and long short-term memory based reduced order surrogate for minimal turbulent channel flow. *Physics of Fluids*, 33(2), 2021.

Oommen, V., Bora, A., Zhang, Z., and Karniadakis, G. E. Integrating neural operators with diffusion models improves spectral representation in turbulence modeling. *arXiv preprint arXiv:2409.08477*, 2024. URL https://arxiv.org/abs/2409.08477.

Parikh, M. H., Fan, X., and Wang, J.-X. Conditional flow matching for generative modelling of near-wall turbulence with quantified uncertainty. *Journal of Fluid Mechanics*, 1029:A32, 2026. doi: 10.1017/jfm.2026.11193.

Park, J. and Choi, H. Toward neural-network-based large eddy simulation: application to turbulent channel flow. *Journal of Fluid Mechanics*, 914:A16, 2021. doi: 10.1017/jfm.2020.931.

Rafiq, K. and Nair, A. G. Cluster-based latent control of unsteady fluid flows. In *Division of Fluid Dynamics Annual Meeting 2025*. APS, 2025.

Rafiq, K., Liao, W., and Nair, A. G. Single-shot prediction of parametric partial differential equations. *arXiv preprint arXiv:2505.09063*, 2025. URL https://arxiv.org/abs/2505.09063.

Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., and Courville, A. On the spectral bias of neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pp. 5301–5310. PMLR, 2019. URL https://arxiv.org/abs/1806.08734.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF*

Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10684–10695, 2022. doi: 10.1109/CVPR52688.2022.01042.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Shaul, N., Chen, R. T., Nickel, M., Le, M., and Lipman, Y. On kinetic optimal probability paths for generative models. In *International Conference on Machine Learning*, pp. 30883–30907. PMLR, 2023.

Shu, D., Li, Z., and Farimani, A. B. A physics-informed diffusion model for high-fidelity flow field reconstruction. *Journal of Computational Physics*, 478:111972, 2023. doi: 10.1016/j.jcp.2023.111972.

Sirignano, J., MacArt, J. F., and Freund, J. B. DPM: A deep learning PDE augmentation method with application to large-eddy simulation. *Journal of Computational Physics*, 423:109811, 2020. doi: 10.1016/j.jcp.2020.109811.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2021.

Whittaker, T., Janik, R. A., and Oz, Y. Turbulence scaling from deep learning diffusion generative models. *Journal of Computational Physics*, 514:113239, 2024. doi: 10.1016/j.jcp.2024.113239.

Wiewel, S., Becher, M., and Thuerey, N. Latent space physics: Towards learning the temporal evolution of fluid flow. *Computer Graphics Forum*, 38(2):71–82, 2019. doi: 10.1111/cgf.13620.

Wu, Y. and He, K. Group normalization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19, 2018. doi: 10.1007/978-3-030-01261-8_1.

Yeh, C.-A., Gopalakrishnan Meena, M., and Taira, K. Network broadcast analysis and control of turbulent flows. *Journal of Fluid Mechanics*, 910:A15, 2021. doi: 10.1017/jfm.2020.965.

Yousif, M. Z., Yu, L., Hoyas, S., Vinuesa, R., and Lim, H. A deep-learning approach for reconstructing 3D turbulent flows from 2D observation data. *Scientific Reports*, 13 (1):2529, 2023. doi: 10.1038/s41598-023-29525-9.

A. Architecture Details

The Stage 1 VAE architecture is shown in Figure 9. The Stage 2 latent flow matching U-Net is shown in Figure 10. A larger latent of size $8 \times 32 \times 32$ (a weaker $16\times$ compression) was also tested; several latent channels exhibited near-zero empirical variance after training, indicating representational redundancy without performance benefit.

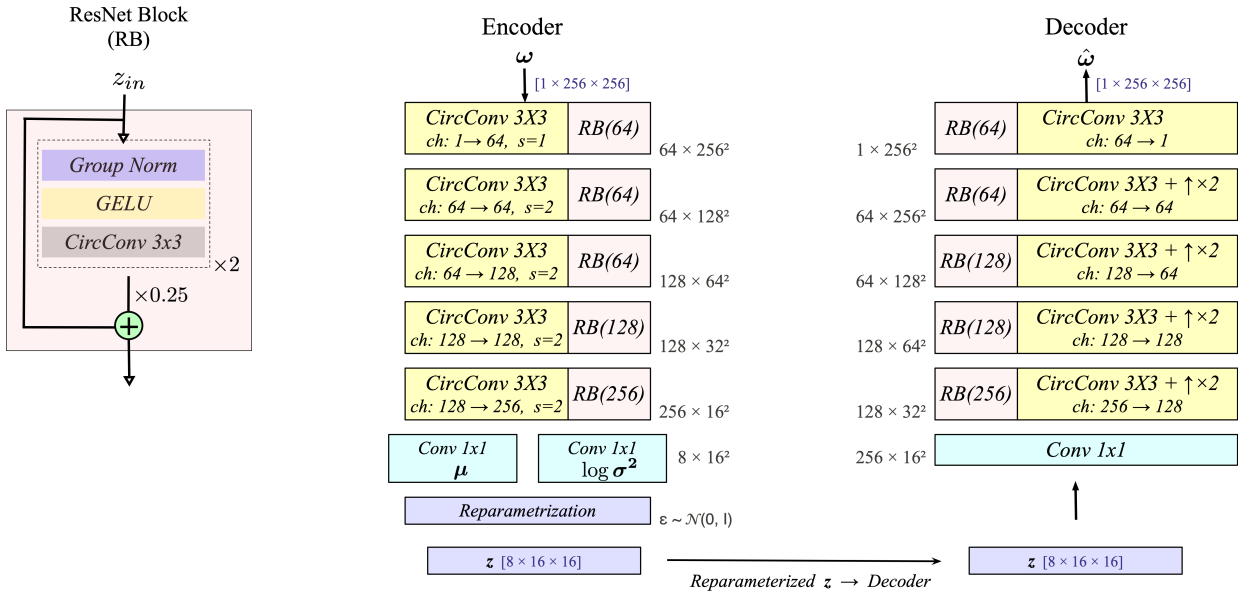

Figure 9. Residual VAE architecture used in Stage 1. Models A and B share the architecture exactly and differ only in the training objective.

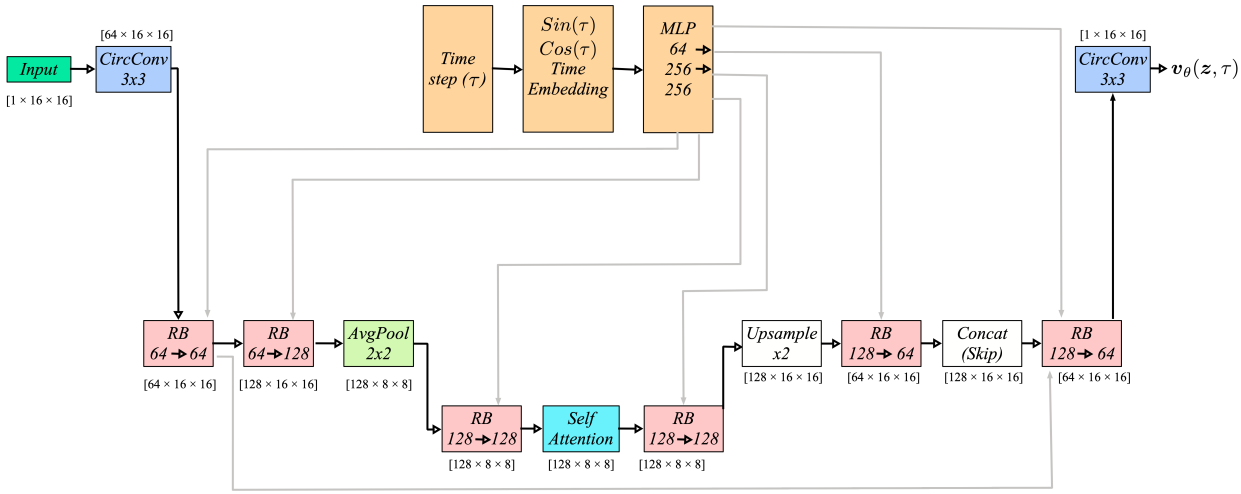

Figure 10. U-Net parameterization of the latent flow matching velocity field $v_\theta(z, \tau)$.

B. Additional Stage 1 Diagnostics

Figure 11 compares a representative held-out field with its Model A and Model B reconstructions and the corresponding DD-band band-pass fields. Figure 12 shows the ensemble distribution of DD-band pointwise MSE underlying the suppression-vs-recovery analysis of Section 5.2.

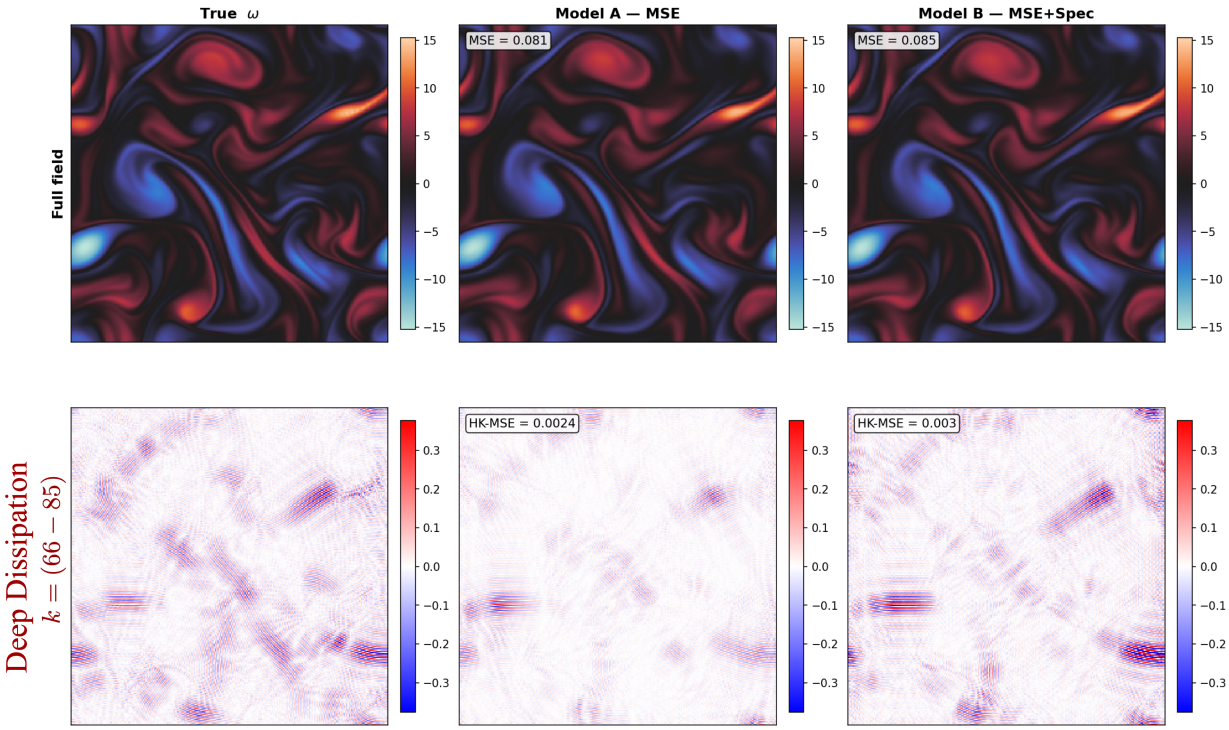

Figure 11. Qualitative reconstruction comparison. Top: full field. Bottom: DD-band ($k = 66$–85) with color scale locked to the true field.

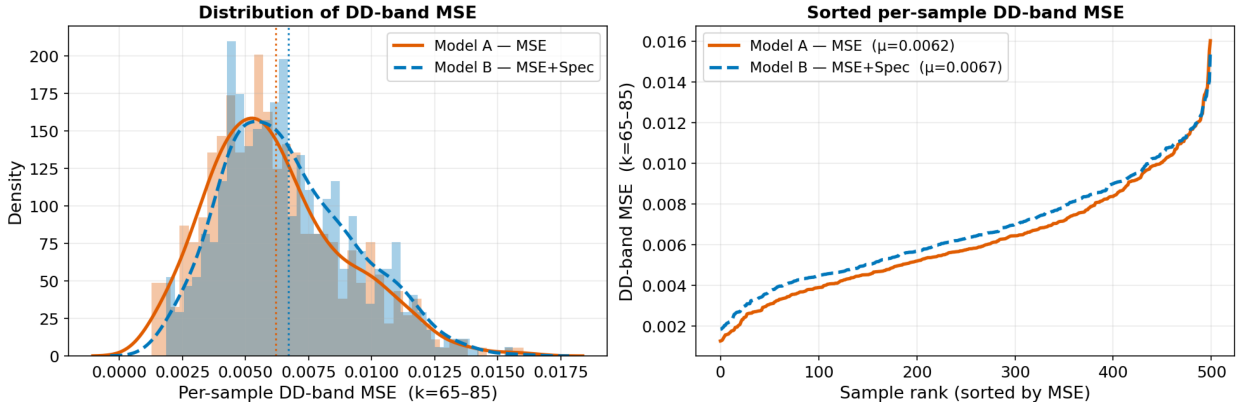

Figure 12. Ensemble DD-band pointwise MSE distribution and sorted per-sample curves. Model B's slightly larger MSE is systematic across the ensemble, not driven by outliers; the support–amplitude decomposition in Figure 4 explains why.