Learning the Entropic Relaxation: Stable Coarse-Grid Emulation with Neural Discrete Equilibrium

Jose Antonio Lara Benitez*, **Kareem Hegazy***, Anastasis Kratsios, Ivan Dokmanić, Maarten V. de Hoop, Michael W. Mahoney

*Equal contribution · NeurIPS 2026

Coarse grids remove the channel that dissipates energy

$$\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla)\mathbf{u} = -\nabla p + \nu \nabla^2 \mathbf{u} \quad | \quad \nabla \cdot \mathbf{u} = 0$$

Incompressible Navier–Stokes

- Dissipation term highlighted in red.
- $\mathbf{u}$ = velocity | p = pressure | ν = kinematic viscosity.

- Viscosity removes features at a dissipation scale L
- Discretizing a system introduces a Nyquist cutoff k_c that disrupts or removes dissipation.
- Fully resolved (stable): grid spacing $\delta x < L$
  - The smallest features are resolved
  - Spectral energy naturally dissipates naturally from high to low frequencies
- Coarse grid (unstable): $\delta x \ll L$
  - The grid cutoff (the Nyquist wavenumber k_c) truncates the energy cascade before it reaches L.
  - Energy piles up at the highest resolved wavenumbers and re-enters as grid-scale noise (aliasing).
  - Learned coarse-grid emulators inherit this, because their training data are already coarsened.

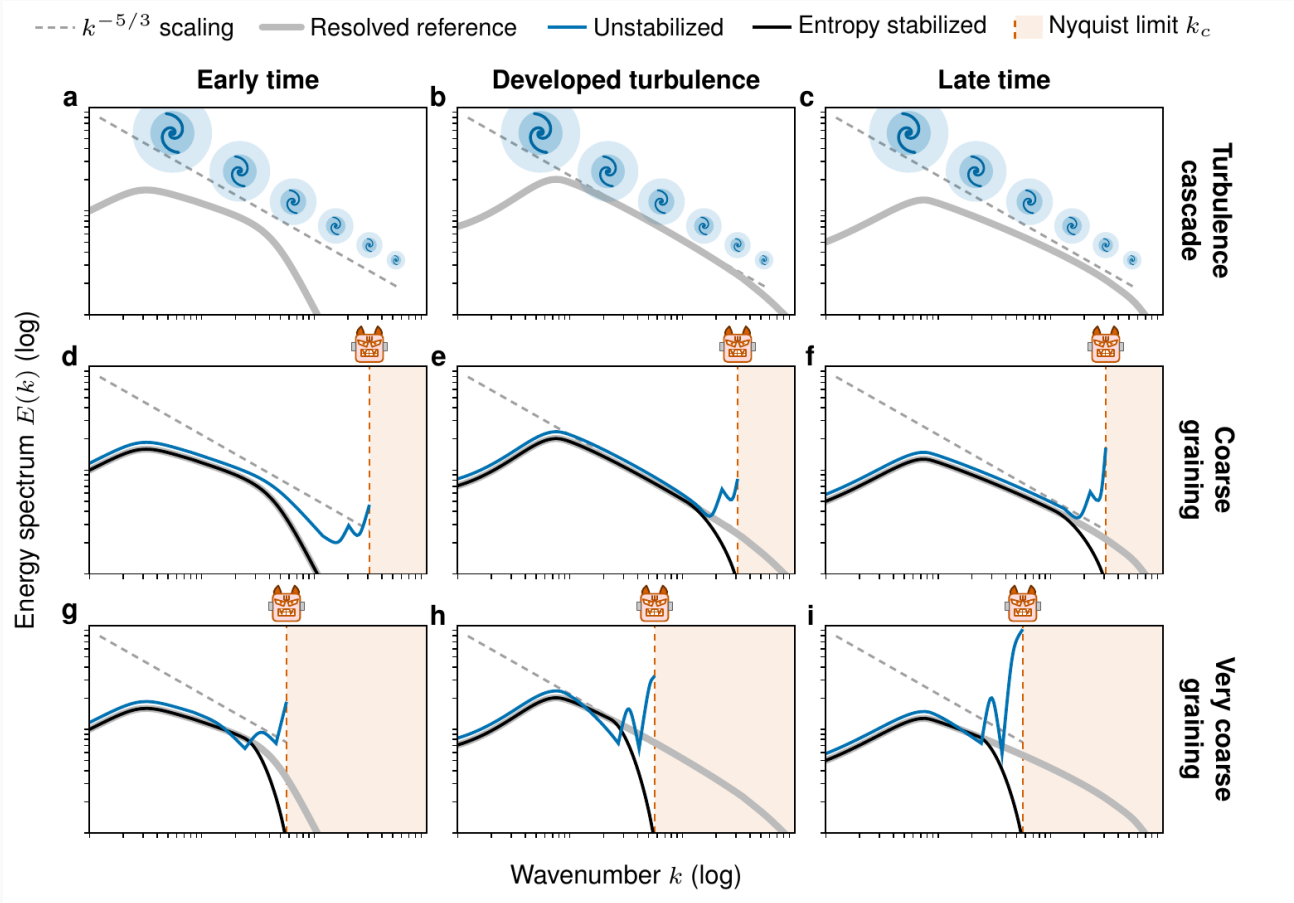


Figure 1: Discretizing a physical system introduces a Shannon Nyquist cutoff at k_c that allows spectral energy to flow only from low to high (Maxwell Demon) while causing spectral energy to buildup. The shaded, blue, and black lines respectively demonstrate expected behavior of simulations at continuous, course, and stabilized evaluations.

Entropy is sensitive to energy loss and gained by discretization

Entropy is a measure of information, order, and disorder

Second Law of Thermodynamics

The total entropy of a closed system cannot decrease: information content can only stay the same or decrease. Dissipation, via viscosity, turns kinetic energy from into thermal energy by dissipating these structures towards an equilibrium. In doing so, entropy increases.

Coarse Grids Remove Information

A coarse grid cannot store subgrid detail, so a correct update must forget it, and forgetting raises entropy. An update that lowers entropy is inventing detail that was never there. That invented detail appears as grid-scale noise (App. A.5).

Energy Buildup Breaks the Second Law

On a coarse grid, energy that should cross the cutoff and turn into heat stays on the grid as small, organized structure. The flow ends up more ordered than physics allows: entropy falls, which violates the second law. The paper calls this an entropy-stability error.

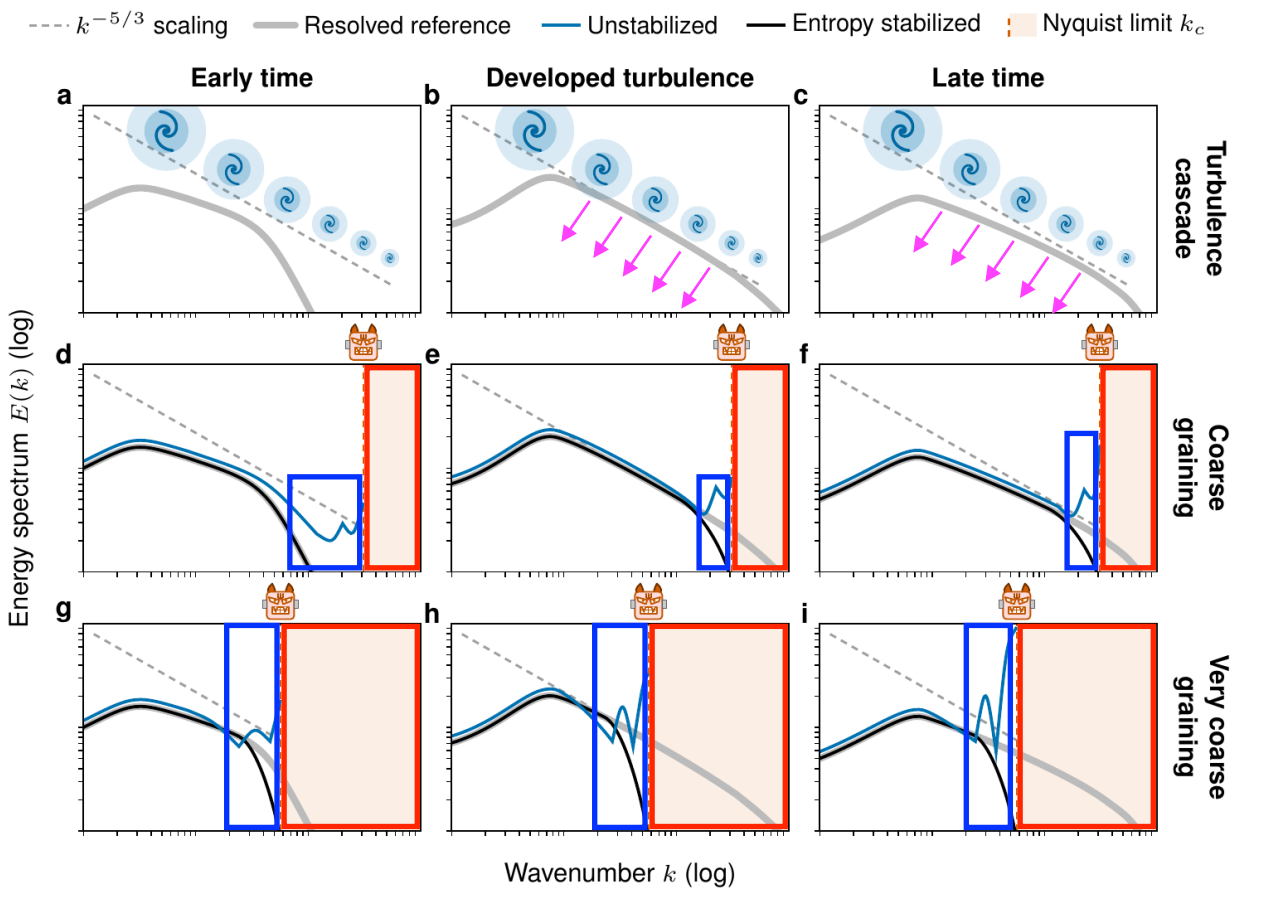


Fig. 1: Discretizing a physical system introduces a Shannon Nyquist cutoff at k_c that allows spectral energy to flow only from low to high (Maxwell Demon) while causing spectral energy to buildup. The shaded, blue, and black lines respectively demonstrate expected behavior of simulations at

NeurDE: a lattice Boltzmann backbone with a learned equilibrium

Kinetic lifting: mesoscopic (particle) scale

$$f_i(t+1, \mathbf{x}+\mathbf{v}_i) = f_i(t, \mathbf{x}) + \frac{1}{\tau}\left[f_i^{\mathrm{eq}}(\mathbf{U}(t, \mathbf{x})) - f_i(t, \mathbf{x})\right]$$

$f_i(t, \mathbf{x})$ is the single particle distribution with with particle velocity v_i

Projection: mesoscopic to macroscopic

$$\mathbf{U} = [\rho, \rho\mathbf{u}, \rho E] = \sum_i [1, v_i, 1/2 v_i^2] f_i$$

The velocity moments over $f_i(t, \mathbf{x})$ give the macroscopic (conserved) observables $\mathbf{U}$.

Collision: nonlinear and local

Unknown/learned physics

$$f_i^{\mathrm{collide}}(t, \mathbf{x}) = f_i(t, \mathbf{x}) + \frac{1}{\tau}\left[f_i^{\mathrm{eq}}(\mathbf{U}(t, \mathbf{x})) - f_i(t, \mathbf{x})\right]$$

Each grid point relaxes toward f_i^{eq}, which includes the system specific closures.

Streaming: linear and nonlocal

$$f_i(t+1, \mathbf{x}+\mathbf{v}_i) = f_i^{\mathrm{collide}}(t, \mathbf{x})$$

Populations shift exactly to the neighboring grid point. Nothing is learned.

Entropically-stabilized NeurDE: swap BGK for the entropic collision

Entropic collision: stability enforcing mechanism

$$f^{\text{collide}} = (1-\beta)f + \beta f^{\text{mirror}}(\alpha) = f + \alpha\beta\left(f^{\text{eq}} - f\right)$$

$f^{\text{collide}}(t, \mathbf{x})$ is modulated between f and f^{eq} by α using the mirror state (Karlin et al. 1999).

Entropic constraints: second law of thermodynamics

$$f^{\text{mirror}}(\alpha) = f + \alpha\left(f^{\text{eq}} - f\right) \qquad H\left(f^{\text{mirror}}(\alpha)\right) = H(f)$$

$f^{\text{mirror}}(t, \mathbf{x})$ has the same entropy as f.

α modulates viscosity to locally impact flow

$$\nu_{\text{eff}}(\alpha) = c_s^2\left(\frac{1}{\alpha\beta} - \frac{1}{2}\right)$$

α	Relaxation rate	Effective viscosity	Local effect
$\alpha = 2$	$\alpha\beta = 1/\tau$	Nominal	Standard BGK
$\alpha < 2$	Slower	Increases	Smoothing: damps high frequencies
$\alpha > 2$	Faster	Decreases	Sharpening: permits backscatter

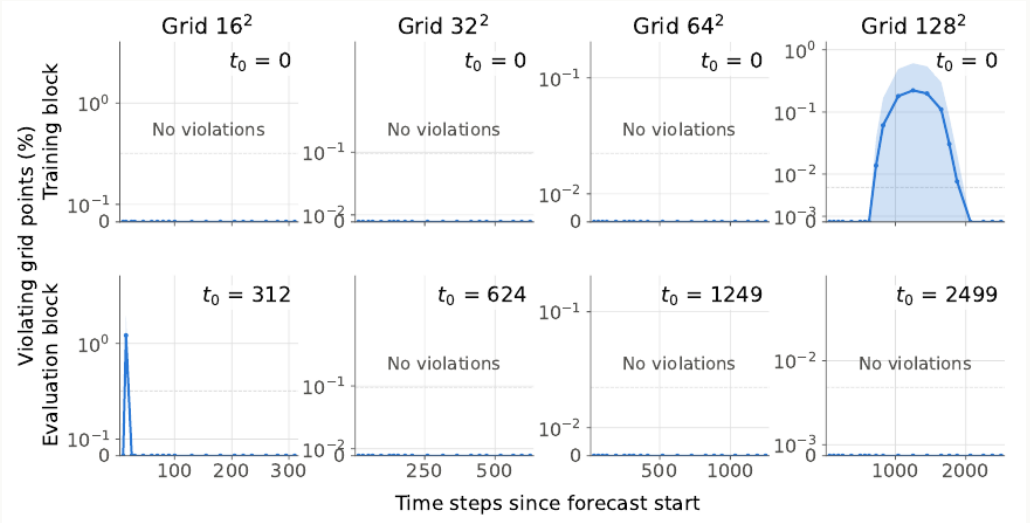


Figure 2: Grid points violating the second law of thermodynamics on the doubly periodic system with $\text{Re} = 50{,}000$.

Karlin, Ferrante & Öttinger, Europhys. Lett. 47, 182 (1999)

SciML baselines fail on the doubly periodic shear layer

- Benchmarking FNO and Walrus on the most resolved (256×256) and easiest experiment: doubly periodic shear layer with $\mathrm{Re} = 50{,}000$.
- FNO (Li et al. 2021) fails within 5 autoregressive steps.
- Walrus (McCabe et al. 2025), a 1.3B-parameter pretrained foundation model, was fine-tuned on the same training data. It is stable to $\Delta t \approx 1250$, but fails to form the vortices by $\Delta t \approx 2500$ on training data, then diverges.

Fourier Neural Operator (FNO)



Figure 3: FNO vorticity rollout on the doubly periodic system Re=50,000.

Fine-tuned Walrus Foundation Model

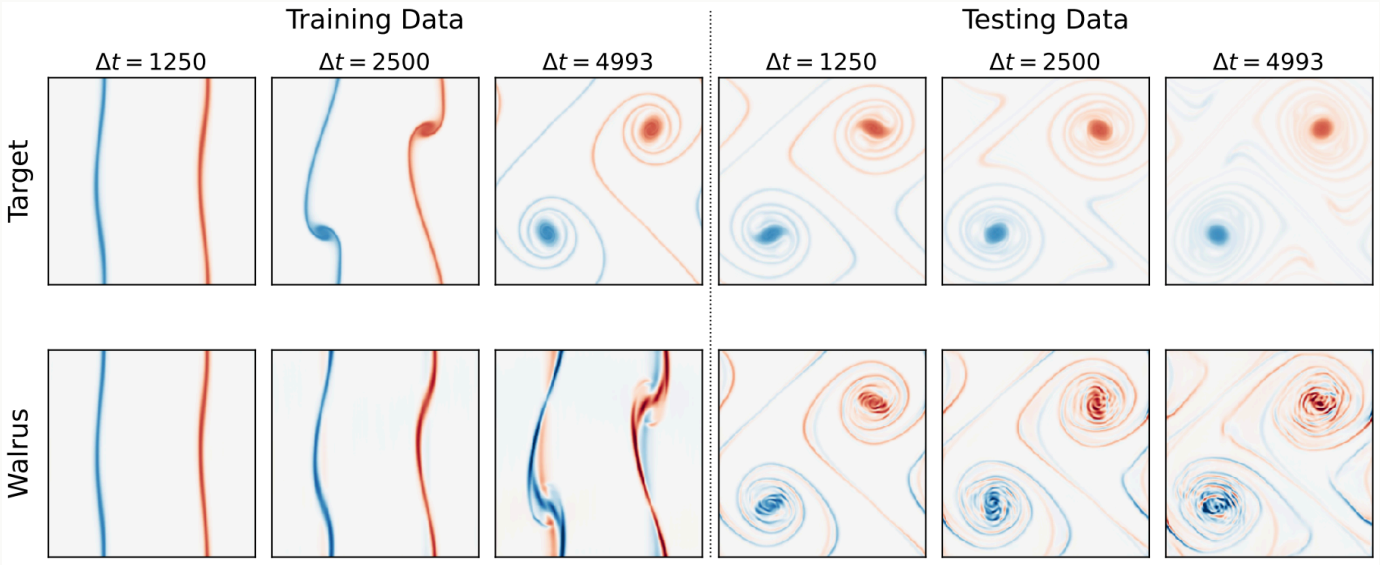


Figure 4: Fine-tuned Walrus foundation model vorticity rollout on the doubly periodic system Re=50,000.

Stability at extremely coarse grids and high Reynolds number

Challenging benchmarks

- Doubly periodic shear layer with $\mathrm{Re} = 100{,}000$
- Compare NeurDE against ELBM, the stabilized numerical simulator that generated the data
- Test on grids from 128×128 down to 16×16

Entropically stabilized NeurDE is stable to 16×16

- Entropically stabilized NeurDE faithfully reproduces the expected features on all tested grids.

Entropic LBM dynamics diverge

- Coarse-grid ELBM shows grid-scale noise by $\Delta t = 625$ on the 64×64 grid.
- ELBM begins breaking down at 128×128 (shown in paper).

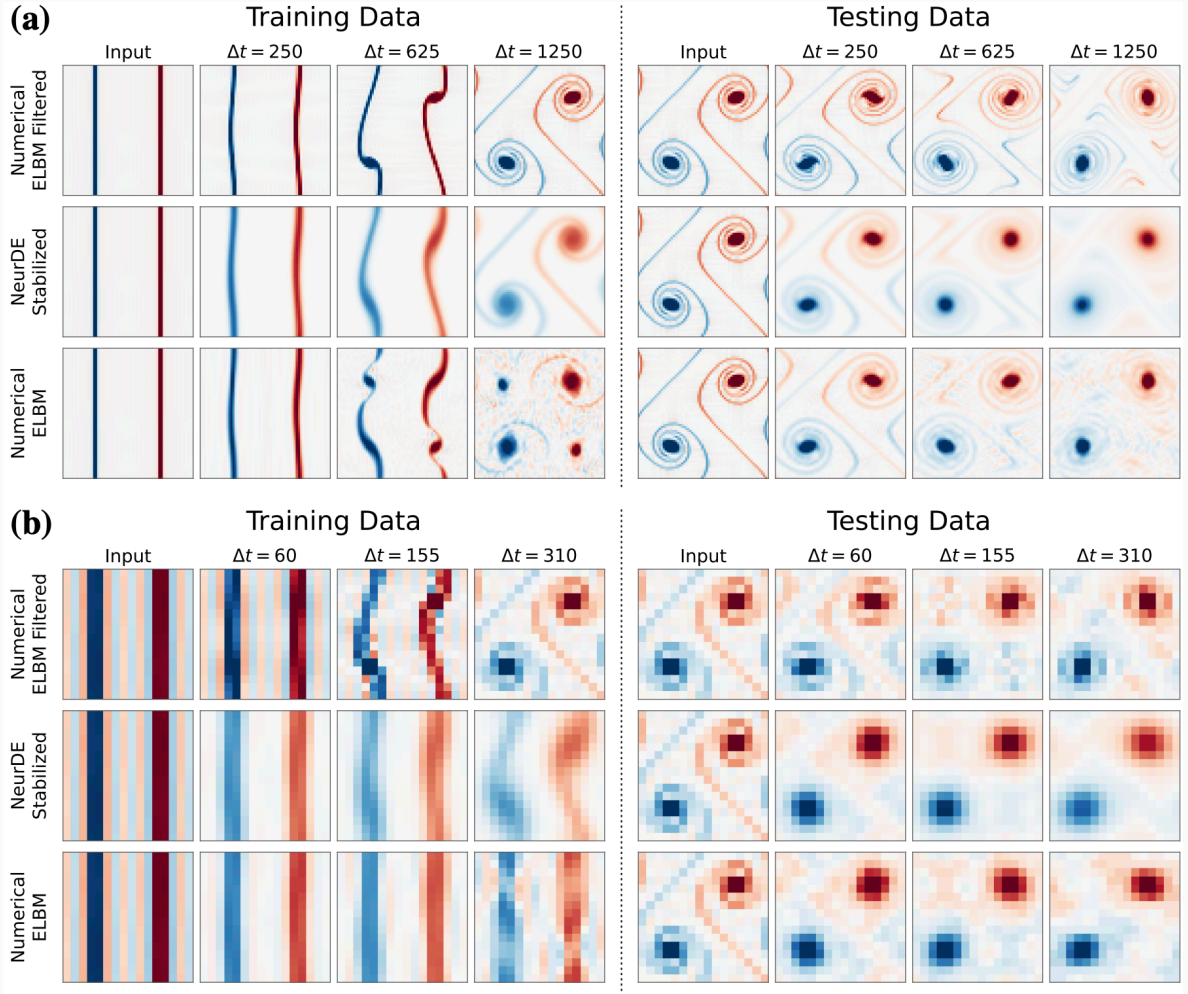


Figure 5: Autoregressive vorticity rollout of stabilized NeurDE (middle rows) and the numerical entropic LBM (bottom row) for coarse grids (a) 64×64 and (b) 16×16.

NeurDE smooths where it needs to and does not otherwise

NeurDE smooths where it must

- The predicted α comparison (Fig. 6) shows that NeurDE captures both the primary and fine grain structures.
- NeurDE also predicts $\alpha > 2$, which increases high-frequency components.

NeurDE does not smooth where it does not need to

- The Taylor-Green vortex experiment demonstrates stable vortices, over smoothing by NeurDE would make the vortices unstable.
- We test NeurDE on the Taylor-Green vortex with $\mathrm{Re} = 50{,}000$.
- NeurDE is not smoothing critical modes for proper dynamics.

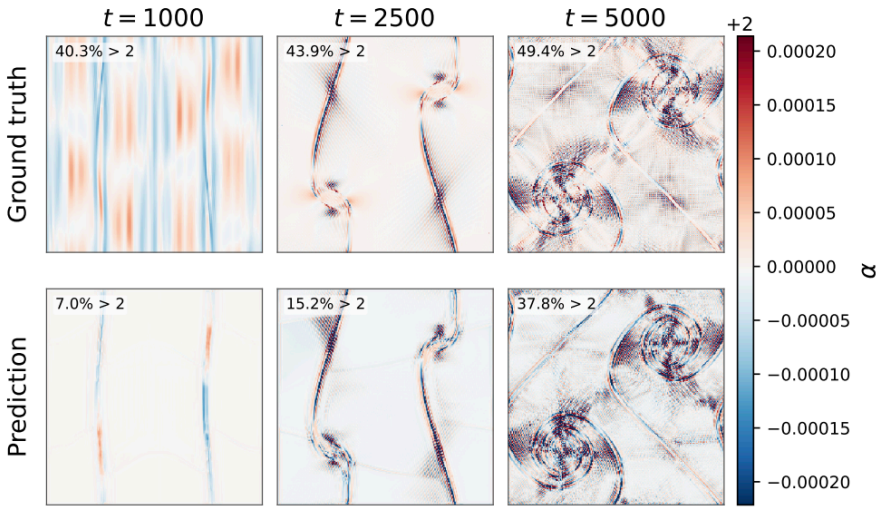


Figure 6: α from ELBM (top) and NeurDE (bottom) on the doubly periodic shear layer $\mathrm{Re} = 50{,}000$.

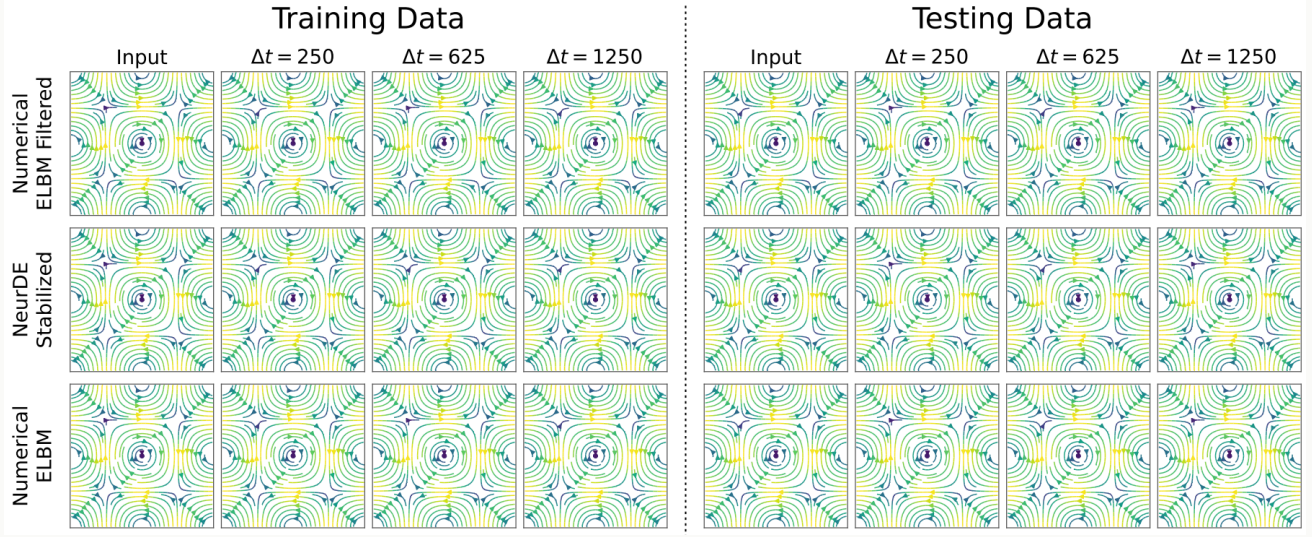


Figure 7: Autoregressive vorticity rollout on the Taylor–Green vortex at $\mathrm{Re} = 50{,}000$ and a 64×64 grid. Rows: filtered reference, NeurDE, coarse ELBM.

Takeaways

- Discretizing a physical systems can alter the physical dynamics being simulated.
- Entropic LBM and entropically stabilized NeurDE offer a means to decrease and increase the high-frequency components near the Nyquist limit.
    - A learned, local α acts as an adaptive viscosity: it smooths where the flow needs it and sharpens where the flow allows it.
- We leverage the second law of thermodynamics to inform how we alter the frequency modes at the Nyquist limit.
- At Re = 100,000, entropically-stabilized NeurDE reproduces the dynamics on grids down to 16 × 16, where the coarse-grid ELBM does not. FNO and Walrus already fail on the easier Re = 50,000 case at 256 × 256.

Code: https://github.com/khegazy/neurde

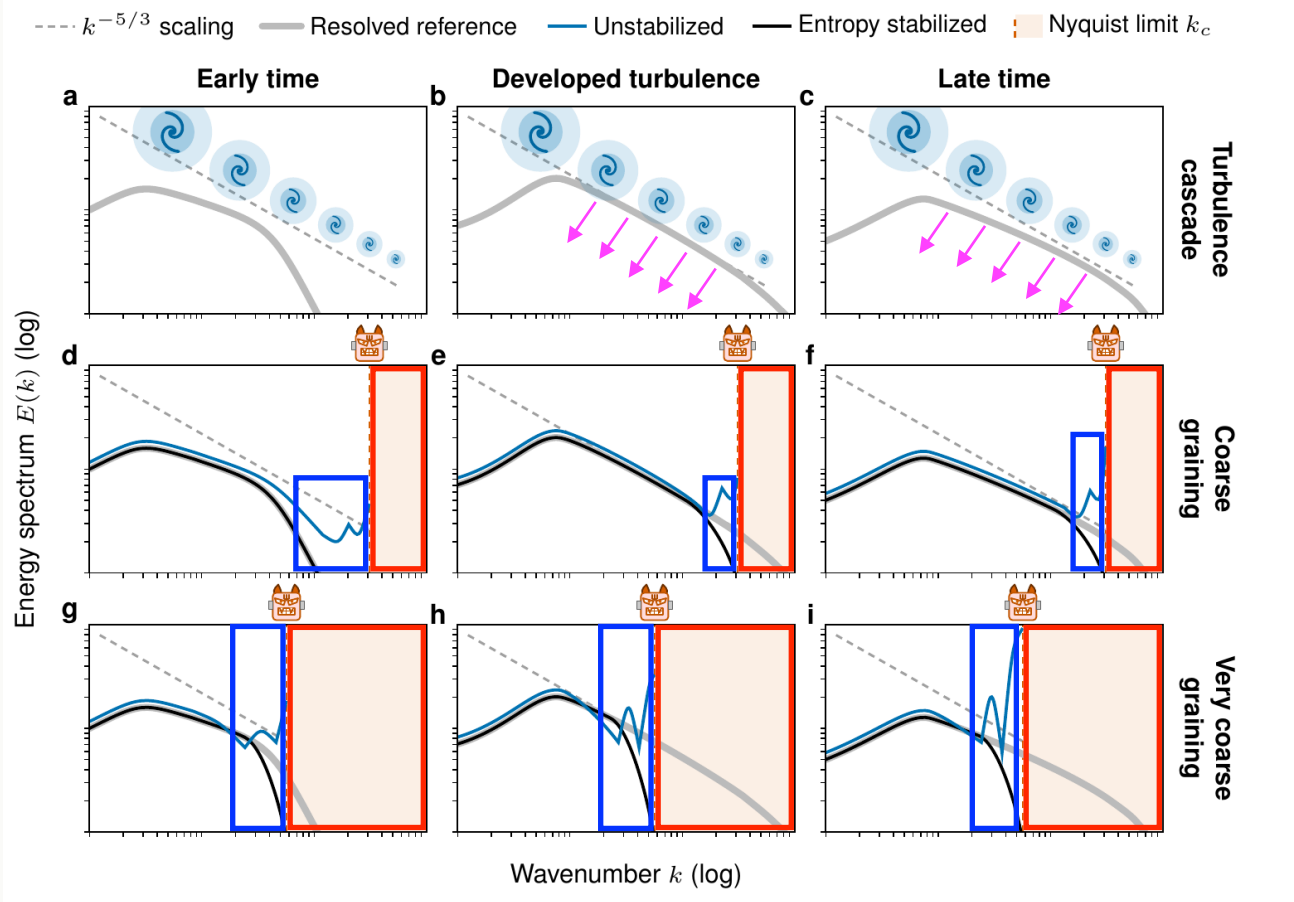


Fig. 1: Discretizing a physical system introduces a Shannon Nyquist cutoff at k_c that allows spectral energy to flow only from low to high (Maxwell Demon) while causing spectral energy to buildup. The shaded, blue, and black lines respectively demonstrate expected behavior of simulations at