LEARNING IN THE TRANSVERSE SUBSPACE: A MINIMAL REPRESENTATION FOR DIVERGENCE-FREE OPERATOR LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Divergence-free vector fields are fundamental state variables in incompressible flows and many PDE systems. Redundant methods, including Neural Conservation Law (NCL) potentials, map multiple auxiliary fields to the same physical divergence-free field. Our experiments show that this redundancy can lower static representation-fitting error. However, their many-to-one, non-invertible mapping does not provide the unique function state required by operator learning, where an evolution operator maps each state function to its future state.

We introduce a minimal representation that encodes a real D-component divergence-free vector field on a D-dimensional domain as a real $(D-1)$-component vector field on the same domain. For periodic fields and closed impermeable fields, the direct transform is invertible, isometric, and angle-preserving. For open nonperiodic flows, we use Fourier extension to construct a compatible periodic field, and a minimum-energy rule selects a reduced representation.

These coordinates provide a direct state space for divergence-free operator learning. Training data are first encoded, and the neural model learns the evolution directly in the $(D-1)$ reduced components. At inference, the predicted reduced field is decoded back to a physical D-component divergence-free field. The model never predicts an ambient D-component velocity and does not need a projection to remove its longitudinal or divergent component. Experiments show that this end-to-end divergence-free formulation achieves lower error and greater robustness. In projection-based and hard-constrained alternatives, a predicted redundant component can be discarded by projection or decoding; because the physical loss does not constrain that direction, it can introduce instability. Our model never represents this longitudinal/null direction.

1 INTRODUCTION

Divergence-free constraints are fundamental geometric structures associated with conservation in physics. In incompressible fluid dynamics, $\nabla \cdot u = 0$ expresses local volume preservation; in electromagnetism and magnetohydrodynamics, $\nabla \cdot B = 0$; and more generally conservation is expressed by $\partial_t \rho + \nabla \cdot j = 0$. Thus, divergence-free structure is a condition of physical admissibility in fluid dynamics, electromagnetism, magnetohydrodynamics, and related conservation-law systems (Chorin, 1968; Tóth, 2000; Richter-Powell et al., 2022).

This structure poses a basic question for machine learning. Neural models are increasingly used to reconstruct physical vector fields and learn their dynamics. Operator learning seeks a function-to-function evolution map

$$\mathcal{G} : u(\cdot, t) \longmapsto u(\cdot, t + \Delta t). \tag{1}$$

An unconstrained neural output does not generally remain divergence-free, so many methods enforce the constraint exactly rather than merely penalizing its violation. The main approaches either use an auxiliary potential or predict in the ambient space and remove the inadmissible component (Kim et al., 2019; Richter-Powell et al., 2022; Tompson et al., 2017).

Potential-based approaches exploit identities such as $\nabla \cdot (\nabla \times A) = 0$ and their differential-form generalizations, so a fixed differential operator maps a learned auxiliary object to a divergence-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

free field. This principle underlies stream-function, vector-potential, and Neural Conservation Law (NCL) constructions (Kim et al., 2019; Richter-Powell et al., 2022). Although they guarantee admissibility, their coordinates are generally redundant and non-identifiable: a 3D divergence-free Fourier mode has two transverse degrees of freedom, whereas a vector potential has three components and a gauge direction. More generally, antisymmetric potentials may use $D(D-1)/2$ components to represent only $D-1$ transverse degrees of freedom. Such redundancy can ease static fitting but does not provide a unique physical-state coordinate.

Thus, admissibility alone is insufficient for operator learning: a state coordinate should be identifiable so that the same physical field yields a unique future state.

Projection-based approaches instead predict an ambient D-component field $\widetilde{u}$ and remove its longitudinal component through a Leray–Helmholtz projection or pressure correction (Chorin, 1968; Tompson et al., 2017). On a periodic Fourier domain,

$$P_k = I - \frac{kk^\top}{|k|^2}, \qquad P_k\big(\widehat{\widetilde{u}}(k) + \alpha k\big) = P_k\widehat{\widetilde{u}}(k) \quad \text{for any } \alpha. \tag{2}$$

the projected loss is insensitive to the longitudinal component generated before projection. The network may therefore spend capacity on a component that is discarded, while that null direction drifts without changing the projected state. This extra freedom can degrade conditioning or rollout robustness.

Pressure-correction schemes expose the same distinction from another angle: the learned quantity is the component to remove, and the admissible field is recovered only after a fixed correction. Such methods can be effective, but they still learn through an ambient or redundant variable before obtaining the physical state.

These observations motivate a simple question: *if a divergence-free state has only $D-1$ intrinsic degrees of freedom, why learn D or more quantities and remove the redundancy afterward?* The Fourier constraint $k^\top \widehat{u}(k) = 0$ places every nonzero coefficient of a D-dimensional divergence-free field in the $(D-1)$-dimensional transverse subspace

$$k^\perp = \{v \in \mathbb{C}^D : k^\top v = 0\}. \tag{3}$$

This reduction is classical in spectral descriptions of incompressible turbulence, where transverse and helical bases represent a three-component mode with two amplitudes (Waleffe, 1992; Kimura & Herring, 2012). Our key observation is to use these intrinsic coordinates as the state variables learned by a neural operator.

We construct a fixed analytic Fourier–Householder transformation that maps a real D-component divergence-free field to a real $(D-1)$-component field on the same domain. At each nonzero mode, a deterministic orthonormal basis $E(k) \in \mathbb{R}^{D\times(D-1)}$ spans $k^\perp$, giving $\widehat{u}(k) = E(k)\widehat{a}(k)$ and $\widehat{a}(k) = E(k)^\top \widehat{u}(k)$; the zero mode is handled separately. For periodic and closed impermeable fields, the transform is invertible, isometric, and angle-preserving. For open nonperiodic flows, Fourier extension and a minimum-energy rule provide a unique reduced representation, although this extension/restriction is generally not isometric (Huybrechs, 2010).

We encode each training trajectory once and learn the reduced evolution $\mathcal{G}_\perp : a(\cdot, t) \mapsto a(\cdot, t + \Delta t)$ directly on the $(D-1)$ transverse components. At inference, a deterministic decoder maps the prediction to a physical divergence-free field. The model therefore never represents the longitudinal degree of freedom and learns the evolution directly in intrinsic coordinates.

Contributions. We introduce a minimal, deterministic $D \to D-1$ representation for real divergence-free fields, with an invertible, isometric transform for periodic and closed impermeable domains and a Fourier-extension rule for open domains. We use these reduced coordinates as the state space of operator learning, eliminating the longitudinal/null direction; experiments show improved accuracy and rollout robustness relative to redundant alternatives.

2 Related Work

Classical spectral treatments of 3D incompressible turbulence already exploit the $3 \to 2$ transverse reduction. Craya–Herring and helical decompositions express each divergence-free Fourier velocity

mode through two transverse amplitudes (Kimura & Herring, 2012; Waleffe, 1992), but they do not provide the arbitrary-D, reversible $D \to D-1$ coordinate layer considered here.

These classical bases provide analytical transverse coordinates, but not a general reversible state map for arbitrary-dimensional neural-operator evolution.

Hard divergence-free constructions in learning systems largely use potentials or projections. Potential-based methods predict an auxiliary scalar, vector, or antisymmetric field and apply a fixed differential map whose range is divergence-free. Deep Fluids, vector-potential coarse-graining, a-Net, Neural Conservation Laws, clawNO, and related models instantiate this principle (Kim et al., 2019; Mohan et al., 2023; Wandel et al., 2021; Richter-Powell et al., 2022; Liu et al., 2024; Tretiak et al., 2022; Cheng et al., 2026; Bisbal et al., 2026). These methods guarantee admissibility, but their auxiliary representation is generally redundant: a 3D vector potential has a longitudinal gauge direction, while an antisymmetric $D \times D$ potential uses $D(D-1)/2$ components to represent only $D-1$ transverse degrees of freedom. Leon and Scheinker (Leon & Scheinker, 2024) exploit the transverse subspace of divergence-free fields for gauge-free electromagnetic vector-field reconstruction, but their same-time map is not a temporal operator-learning map.

Exact projection methods predict an ambient field and apply a Leray/Helmholtz projection, as in the physics-constrained latent Neural ODE, hard-constraint operator framework, and deterministic *Project* branch (Shankar et al., 2022; Duruisseaux et al., 2024; Li et al., 2026); the network may still produce a longitudinal component that is discarded. A related CFD-inspired route, used by the Data-Driven Projection Method and FluidNet, instead learns the pressure/longitudinal correction to subtract from a provisional velocity rather than a minimal transverse coordinate or an exact Fourier projector (Yang et al., 2016; Tompson et al., 2017).

Kernel- and prior-based approaches construct divergence-free function spaces directly: DFK and Fluids You Can Trust use divergence-free kernels or bases for flow reconstruction and operator learning, while Linearly Constrained Gaussian Processes and Gaussian Process Hydrodynamics impose differential constraints through Gaussian-process priors (Ni et al., 2025; Sharma et al., 2026; Jidling et al., 2017; Owhadi, 2023). Across these tasks, the common mechanism is an analytically constrained function space rather than a reversible minimal coordinate transform.

Our approach instead parameterizes each nonzero Fourier mode directly by its intrinsic $D-1$ coordinates in $k^\perp$. After the stated zero-mode convention, the map is reversible and contains no longitudinal null direction, separating the geometry of incompressibility from the learning architecture and dynamics.

3 METHODOLOGY

We first show that a real D-component divergence-free vector field occupies only an intrinsic $(D-1)$-dimensional transverse subspace at every nonzero Fourier mode. Consequently, the field admits an equivalent $(D-1)$-component representation on the same spatial domain. Learning the evolution or operator directly in these intrinsic coordinates is therefore sufficient: decoding any reduced state recovers a D-component divergence-free field by construction. For $u : \Omega \subset \mathbb{R}^D \to \mathbb{R}^D$, define $\widehat{u}(k) = \int_\Omega u(x) e^{-ik\cdot x}\, dx$. Starting from the product rule,

$$\nabla\cdot\big(ue^{-ik\cdot x}\big) = (\nabla\cdot u)e^{-ik\cdot x} + u\cdot\nabla e^{-ik\cdot x} = (\nabla\cdot u)e^{-ik\cdot x} - i(k^\top u)e^{-ik\cdot x}, \tag{4}$$

$$\int_\Omega \nabla\cdot\big(ue^{-ik\cdot x}\big)\, dx = \int_{\partial\Omega} (u\cdot n)e^{-ik\cdot x}\, dS = \int_\Omega (\nabla\cdot u)e^{-ik\cdot x}\, dx - ik^\top \widehat{u}(k), \tag{5}$$

$$\therefore \qquad k^\top \widehat{u}(k) = i\int_{\partial\Omega} (u\cdot n)e^{-ik\cdot x}\, dS - i\int_\Omega (\nabla\cdot u)e^{-ik\cdot x}\, dx. \tag{6}$$

Thus, under the two conditions $\nabla\cdot u = 0$ in Ω and $\int_{\partial\Omega}(u\cdot n)e^{-ik\cdot x}\, dS = 0$, we obtain exactly $k^\top\widehat{u}(k) = 0$. The second condition holds for impermeable boundaries because $u\cdot n = 0$ pointwise, and for periodic boundaries because opposite-face contributions cancel for the periodic Fourier modes. Hence, for every $k \neq 0$, $\widehat{u}(k) \in k^\perp$, so each D-component Fourier mode contains exactly $D-1$ independent degrees of freedom, in the setting where the above conditions are satisfied. For general non-periodic boundary conditions, the same result can be obtained through Fourier extension to a periodic computational domain, after which the same transverse construction applies.

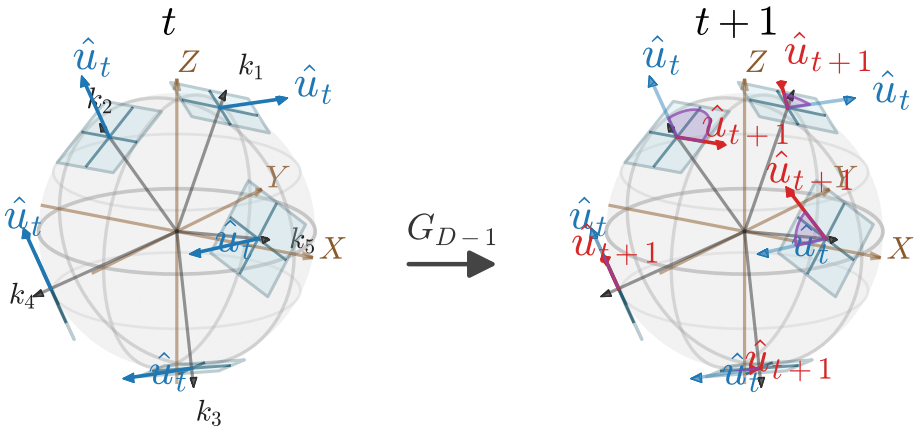


Figure 1: Mode-wise function evolution in the minimal transverse representation. For every nonzero wavenumber k_i, both the input mode $\widehat{u}_t(k_i)$ and the decoded next mode $\widehat{u}_{t+1}(k_i)$ lie in $k_i^\perp$. The learned map $\mathcal{G}_{D-1}$ evolves only the $D-1$ intrinsic transverse coordinates.

This observation suggests learning the dynamics directly in the transverse subspace: every nonzero velocity mode lies and evolves in $k^\perp$, a $(D-1)$-dimensional space. Projection-and-generate methods predict a D-component field before projecting it into $k^\perp$, while potential-based constructions learn redundant vector or matrix coordinates before decoding them to the same $D-1$ degrees of freedom. We instead learn the $D-1 \to D-1$ operator directly in $k^\perp$.

Accordingly, for each nonzero pair $\{k, -k\}$, we choose a canonical representative $\kappa(k) = \kappa(-k)$ and construct an orthogonal matrix $Q(k)$ whose last column is $\kappa/|\kappa|$; the remaining columns $E(k) \in \mathbb{R}^{D\times(D-1)}$ form an orthonormal basis of $k^\perp$. A convenient choice is the Householder matrix

$$Q(k) = I - 2\frac{v_k v_k^\top}{v_k^\top v_k}, \qquad v_k = e_D - \frac{\kappa}{|\kappa|}. \tag{7}$$

The degenerate aligned case means $\kappa/|\kappa| = e_D$, for which $v_k = 0$ and the Householder formula has a zero denominator; in this case we set $Q(k) = I$. This gives the unique coordinates $\widehat{u}(k) = E(k)\widehat{a}(k)$ and $\widehat{a}(k) = E(k)^\top \widehat{u}(k) \in \mathbb{C}^{D-1}$. The Householder frame is inexpensive, depends only on k, and lets us learn only the evolution of these $D-1$ transverse components. Since a standard Fourier neural operator takes real physical-space fields as input and output while transforming internally to Fourier space, we inverse-transform $\widehat{a}$ and learn on a real $(D-1)$-channel field $a(x)$ over the same D-dimensional domain. Enforcing the paired convention $Q(-k) = Q(k)$, hence $E(-k) = E(k)$, preserves conjugate symmetry: for real u, $\widehat{u}(-k) = \overline{\widehat{u}(k)}$ implies $\widehat{a}(-k) = \overline{\widehat{a}(k)}$, so $a(x)$ is real; decoding reverses the map by FFT, multiplication by $E(k)$, and inverse FFT. The zero mode is handled separately.

Three realizations. We use three cases, with M_D denoting restriction to the physical domain, E the block-diagonal transverse basis, and J_K injection of retained Fourier modes. Case I applies directly to periodic or closed impermeable data:

$$A_0 = F_D^{-1} E F_{D-1}. \tag{8}$$

Case II handles open nonperiodic data by Fourier extension before the transverse transform and restriction:

$$A_E = M_D \widetilde{F}_D^{-1} \widetilde{E} \widetilde{F}_{D-1}. \tag{9}$$

Because restriction makes this extension nonunique, its minimum-energy pseudoinverse selects a canonical representation.

Case III also uses an extended domain but retains a conjugate-symmetric mode set K with $|K| = N$:

$$A_K = M_D \widetilde{F}_D^{-1} \widetilde{E} J_K F_{D-1,N}. \tag{10}$$

Thus Cases I and III are direct one-to-one transforms on their admissible spaces, whereas Case II requires canonicalization.

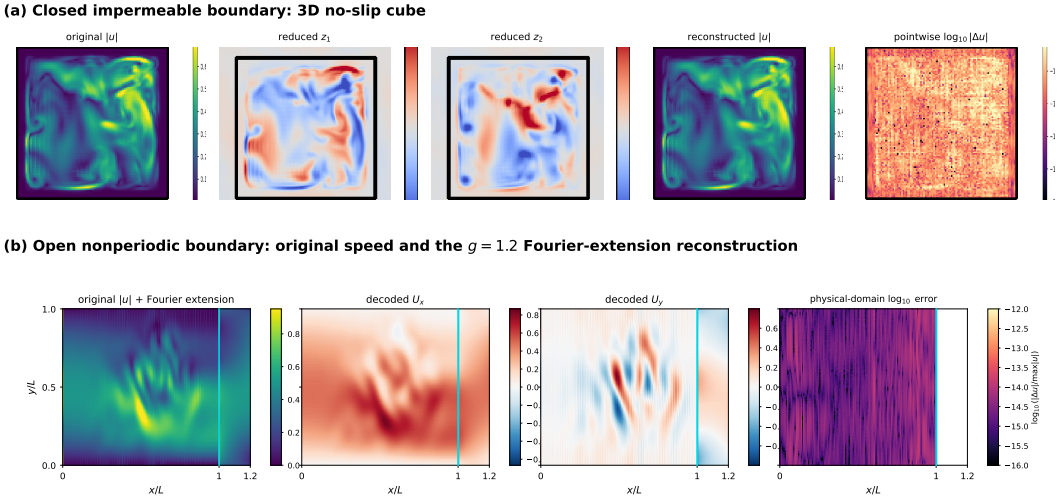


Figure 2: Machine-precision reconstruction with the proposed $D \to D-1 \to D$ transform for a closed 3D domain and an open nonperiodic 2D channel.

Only Case II has a many-to-one ambiguity, removed by the minimum-energy pseudoinverse; retaining sufficient modes is important for stable open-boundary reconstructions.

In practice, thresholding the pseudoinverse suppresses near-null directions; in Case III, full rank of the retained mode map is required for reversibility.

Theorem 1 (Reversible $D \to D-1$ representation)**.** *For Case II, the canonical encoder and decoder are $v = A_E^{\dagger} u$ and $u = A_E v$ on $u \in \mathrm{Range}(A_E)$ and $v \in (\ker A_E)^{\perp}$. They are mutually inverse on these spaces, giving a bijection between admissible divergence-free fields and canonical $(D-1)$-component representations; the zero Fourier mode is stored separately.*

Unlike the redundant raw parameterizations used by NCL, Theorem 1 leaves no unresolved gauge/null direction after the stated restriction. The direct transform A_0 additionally preserves inner products, distances, angles, L^2 and H^s norms, spatial-derivative norms, and time-derivative norms, because the Fourier transforms are unitary and $E(k)$ has orthonormal columns inherited from the orthogonal Householder matrix; it commutes with fixed spatial and temporal differentiation, carries discrete and continuous dynamics by an exact coordinate conjugacy, and preserves Jacobian spectra, singular values, local Euclidean conditioning, and phase-space divergence. Cases II and III retain exact invertibility under their stated canonical/full-rank conditions, but restriction and truncation generally remove the standard Euclidean isometry of Case I. Detailed operator dimensions, kernel structure, and proofs are given in Appendix B.

The zero mode deserves separate treatment because incompressibility imposes no constraint at $k = 0$ and the transverse basis is defined only for $k \neq 0$. Thus a nonzero mean velocity is stored as the D-vector $\bar{u} = \widehat{u}(0)$ alongside v; in the common zero-mean setting used by many periodic benchmark flows, this side channel vanishes identically.

The NCL parameterizations (Richter-Powell et al., 2022) make this redundancy explicit. For its antisymmetric form, $M^{\top} = -M$ and $\widehat{u}(k) = i\widehat{M}(k)k$; hence the linear map $T_k : M \mapsto iMk$ sends the $D(D-1)/2$ independent entries of M into the $(D-1)$-dimensional space $k^{\perp}$. Therefore

$$\dim \ker T_k = \frac{D(D-1)}{2} - (D-1) = \frac{(D-1)(D-2)}{2}, \tag{11}$$

so M and $M + C$ produce the same field for any antisymmetric C satisfying $Ck = 0$. Equivalently, in coordinates with $k = |k| e_D$, $M = \left(\begin{smallmatrix} B & a \\ -a^{\top} & 0 \end{smallmatrix}\right)$ gives $Mk = |k|(a, 0)^{\top}$, and the entire block $B \in \mathfrak{so}(D-1)$ is invisible. NCL's vector form is also non-injective: $M = J_b - J_b^{\top}$ yields $\widehat{u}(k) = -|k|^2 P_k \widehat{b}(k)$, so $\widehat{b}(k)$ and $\widehat{b}(k) + \alpha(k)k$ are identical after decoding. Thus the matrix form uses $D(D-1)/2$ quantities and the vector form uses D, although the physical mode contains

only $D-1$ degrees of freedom. This non-injectivity makes the construction non-invertible and complicates learning one-to-one encoded dynamics in operator learning.

4 EXPERIMENTS

4.1 REPRESENTATION LEARNING

We initially expected the minimal, invertible transverse coordinates to be advantageous for representation learning as well. The experiments do not support this expectation: the NCL potential representations, especially the matrix form, can reach lower loss than the minimal coordinates. A likely reason is that a many-to-one decoder gives each physical divergence-free field many equivalent representations, so finite-step optimization may reach any of many good solutions rather than one unique coordinate. The 10 random-initialization runs directly support this interpretation: our $D-1$-component reduced vector field converges to the same field in every run, whereas NCL's D-component vector potential B and antisymmetric matrix potential A (with $D(D-1)/2$ independent entries) vary across runs even when they decode to the same divergence-free field. Our controlled sweep further confirms that more redundant dimensions can lower the representation error. This advantage does not carry over to operator learning, where a non-invertible code cannot uniquely identify the state to evolve and decode; there an invertible $D-1$ representation defines an unambiguous reduced dynamics. Details are given in Appendix K.

4.2 OPERATOR LEARNING

Benchmarks and evaluation protocol. We evaluate the minimal transverse representation on three divergence-free systems: 2D NS ($2D \to 1D$ velocity), 3D NS ($3D \to 2D$ velocity), and 2D MHD ($4D \to 2D$ for velocity and magnetic field). Across all benchmarks, we compare Raw, Leray, Potential, and Latent2 (ours), with FNO-vorticity added for 2D NS. Models use the same Fourier neural-operator framework and physical-space evaluation; optimizer settings may differ to accommodate the representations' spectral scales, and comparisons use converged validation/holdout error. The reduced Latent2 state also lowers memory use and training cost at matched resolution. For 2D MHD, the two fields are compressed separately and evolved jointly in the two-channel latent state; all predictions are decoded before physical-space evaluation.

2D MHD. In 2D MHD, Latent2 obtains the lowest physical-field error under the common holdout evaluation protocol. The Latent2 model is trained with optimization settings adapted to the scale of its representation, so this comparison reflects the best stably converged model for each representation rather than forcing identical optimizer hyperparameters.

Model	Val. 10-step mean	Holdout 10-step mean	Holdout step-10
A—Raw	0.1343289	0.1350950	0.1847896
B—Leray	0.1332429	0.1338184	0.1917674
C—Potential	0.1005770	0.1012601	0.1465121
D—Latent2 (ours)	**0.0993804**	**0.0996544**	**0.1421250**

Table 1: 2D MHD physical relative-L_2 error. Lower is better.

The gain over the Potential baseline is modest, but the result shows that separately compressing the velocity and magnetic fields to one transverse scalar each does not impair dynamical prediction and can slightly improve the final physical-field error after sufficient training.

3D Navier–Stokes. The 3D NS results show a horizon-dependent trade-off. Raw and Leray are slightly better at very short horizons, whereas Latent2 becomes the lowest-error representation as the closed-loop rollout becomes longer.

Latent2 is not the best one-step predictor, and the models remain close at $h = 5$. At $h = 10$ and $h = 20$, however, Latent2 has the lowest error. This suggests that the principal benefit of a minimal transverse state need not appear in local one-step fitting; it can instead emerge in the stability of repeated autoregressive evolution.

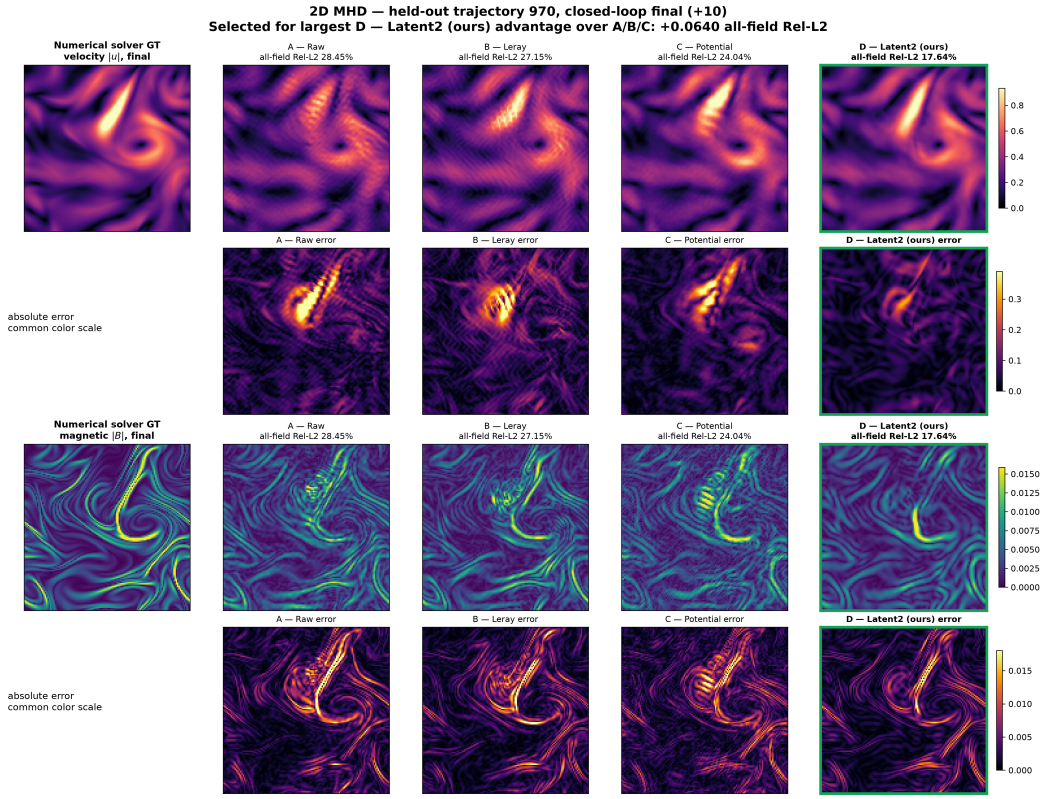


Figure 3: 2D MHD closed-loop prediction on a selected held-out trajectory at the final $+10$ step. The top and middle rows compare velocity and magnetic-field magnitudes, and the lower rows show absolute errors on common color scales. This instance is selected for its large Latent2 advantage over A/B/C; Table 1 remains the aggregate evaluation.

Model	$h = 1$	$h = 5$	$h = 10$	$h = 20$
A—Raw	**0.074044**	0.297128	0.643126	1.060287
B—Leray	0.075585	**0.297055**	0.617995	1.010129
C—Potential	0.078994	0.308620	0.626087	1.079068
D—Latent2 (ours)	0.082842	0.308565	**0.592485**	**1.000849**

Table 2: 3D NS physical relative-L_2 error at different rollout horizons. Lower is better.

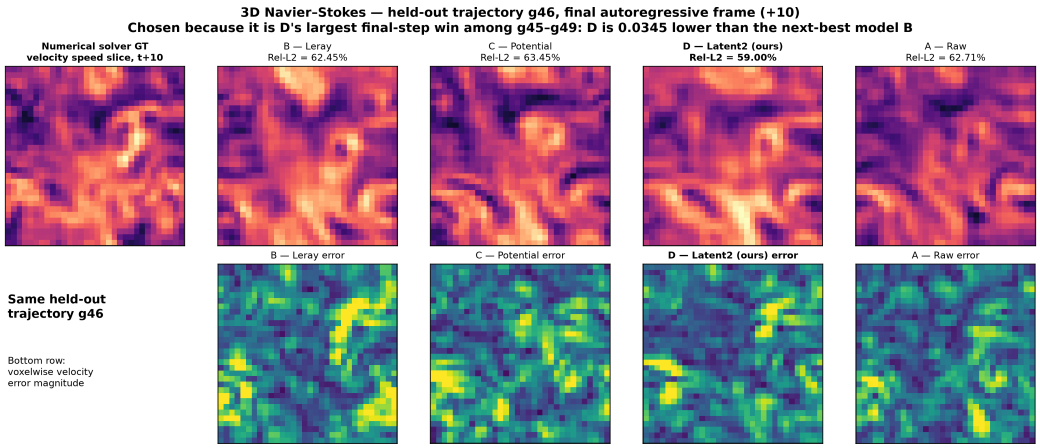


Figure 4: 3D Navier–Stokes final autoregressive $+10$ frame for held-out trajectory g46. The panels show the solver velocity-speed slice, predictions from the four representations, and their voxelwise velocity-error magnitudes. This selected trajectory illustrates the long-horizon advantage; Table 2 reports the aggregate horizon-wise result.

2D Navier–Stokes. For 2D NS, we additionally compare against FNO-vorticity. All five models are reevaluated on the same held-out initial conditions with the same input history, numerical-solver target, and closed-loop rollout, and are compared after reconstruction to physical velocity.

Model	Clean step-10 velocity Rel-L_2
FNO-vorticity	**0.039315**
D—Latent2 (ours)	**0.075430**
B—Leray	0.436152
A—Raw	0.596742
C—Potential	1.003959

Table 3: 2D NS clean step-10 physical velocity error under the common reevaluation protocol.

FNO-vorticity remains the strongest clean baseline in 2D NS. Latent2 does not surpass it, but it substantially outperforms the Raw, Leray, and Potential models in this evaluation. On an independent 50-trajectory holdout, the 10-step mean physical relative-L_2 error is 0.0516114 for Latent2 and 0.0227320 for FNO-vorticity, confirming the strength of the vorticity coordinate in two dimensions. The two scalar representations are closely related, but they induce systematically different spectral weightings in the training objective. In particular, their losses differ by wavenumber-dependent factors involving $|k|$, so the same physical prediction error can receive different emphasis across Fourier bands depending on whether the model predicts vorticity, potential variables, or the transverse latent. FNO-vorticity is therefore an especially informative baseline because it also reduces the two physical velocity components to a single scalar state while using a different spectral weighting. The exact relations and the corresponding loss formulas are given in Appendix H.

Robustness. We also evaluate solver-integrated adversarial perturbations. In 3D NS, under the same perturbation optimized against the Leray model, Latent2 exhibits a much smaller absolute increase in $h = 10$ error:

Model	Absolute error increase at $h = 10$
B—Leray	0.001056259
D—Latent2 (ours)	**0.000028114**

Table 4: 3D NS response to the same B-targeted observation perturbation.

A related transfer effect appears in 2D NS. For the same perturbation optimized against FNO-vorticity, the full-trajectory attack-loss ratio is $3.7538\times$ on FNO-vorticity but only $1.5182\times$ when the identical perturbation is transferred to Latent2. These controlled cases show that the transverse representation can substantially reduce perturbation amplification for adversarial directions optimized against another representation.

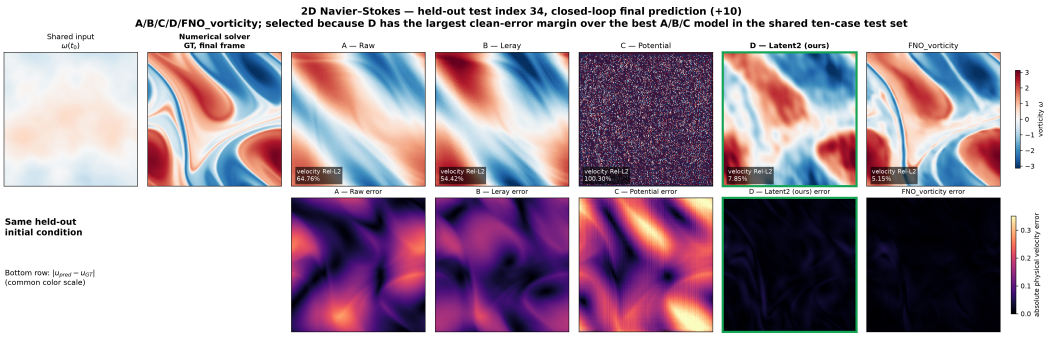


Figure 5: The same 2D Navier–Stokes held-out case with the FNO-vorticity prediction included. At this selected instance, FNO-vorticity has the lowest physical velocity error, followed by Latent2; the C–Potential rollout blows up numerically in this individual case, while the aggregate clean-error comparison is reported in Table 3.

The Leray model also exposes a structural source of possible numerical sensitivity. Its raw output contains both transverse and longitudinal components, but the latter is removed by projection and is therefore only weakly constrained by the physical loss; in our current diagnostic, approximately 7.2–8.0% of the raw response magnitude is discarded by this projection. This null-direction leakage provides a plausible mechanism for extra sensitivity during rollout, whereas Latent2 represents only the transverse degrees of freedom and has no such discarded longitudinal component.

Representation scale and optimization stability. Using the same FNO architecture does not make identical optimizer settings appropriate across representations. Raw, potential, vorticity, and transverse variables describe the same physical state but have different spectral scalings, so changing the prediction variable changes Fourier-band weighting and gradient scales. The exact relations are summarized in Appendix H.

Optimization settings tuned for physical fields may therefore underperform for low-amplitude latent or potential variables. We treat representation and optimization as coupled choices, allowing stable convergence while keeping the data, solver, model capacity, and physical-space evaluation fixed; comparisons are based on converged physical-space validation/holdout error rather than raw losses in differently scaled coordinates.

5 Conclusion

We introduced a minimal $D \to D - 1$ transverse representation that parameterizes divergence-free vector fields directly by their intrinsic Fourier degrees of freedom, yielding an invertible and isometric coordinate system for the periodic/closed settings considered here and eliminating the longitudinal degree of freedom from operator learning. Across 2D MHD, 3D Navier–Stokes, and 2D Navier–Stokes, the representation is competitive with raw, projection-based, and potential-based alternatives, with particularly encouraging long-horizon and controlled-robustness results, while also reducing the number of state channels used during training. At the same time, the empirical improvement is not uniformly large: the construction is mathematically clean, but its practical advantage depends on the problem and optimization setting, and in 2D Navier–Stokes the vorticity formulation remains a stronger clean baseline.

The broader lesson is that divergence-free operator learning benefits from separating physical coordinates from redundant ambient components. Exact transverse coordinates remove the longitudinal null direction before prediction, while the decoder preserves the physical field and enables direct comparison in the original variables. The experiments therefore support the representation as a compact and reversible alternative for divergence-free dynamics, with the choice of coordinates, numerical realization, and optimization treated as one coupled design decision.

AI USE STATEMENT

Generative AI tools were used solely for language editing of the manuscript, including improving grammar, wording, readability, and sentence-level clarity. They were not used to generate experimental results or alter the reported results and conclusions. All AI-assisted edits were reviewed by the author, who takes full responsibility for the final content of the manuscript.

REFERENCES

Javier Bisbal, Julio Sotelo, Hernán Mella, Oliver Welin Odeback, Joaquín Mura, David Marlevi, Junya Matsuda, Kotomi Iwata, Tetsuro Sekine, Cristian Tejos, and Sergio Uribe. Unsupervised 4D Flow MRI Velocity Enhancement and Unwrapping Using Divergence-Free Neural Networks. arXiv:2604.00205, 2026.

Raoul Bott and John W. Milnor. On the Parallelizability of the Spheres. *Bulletin of the American Mathematical Society*, 64(3):87–89, 1958.

Jinbao Cheng, Jianguo Huang, Haoqin Wang, and Tao Zhou. Decoupled Divergence-Free Neural Networks Basis Method for Incompressible Fluid Problems. arXiv:2603.17906, 2026.

Alexandre J. Chorin. Numerical solution of the Navier–Stokes equations. *Mathematics of Computation*, 22(104):745–762, 1968.

Valentin Duruisseaux, Miguel Liu-Schiaffini, Julius Berner, and Anima Anandkumar. Towards Enforcing Hard Physics Constraints in Operator Learning Frameworks. ICML AI for Science Workshop, 2024.

Daan Huybrechs. On the Fourier extension of nonperiodic functions. *SIAM Journal on Numerical Analysis*, 47(6):4326–4355, 2010.

Yuchi Jiang, Zhijie Li, Yunpeng Wang, Huiyu Yang, and Jianchun Wang. An Implicit Adaptive Fourier Neural Operator for Long-term Predictions of Three-dimensional Turbulence. *Acta Mechanica Sinica*, 42(6), 2026. doi:10.1007/s10409-025-25478-x.

Carl Jidling, Niklas Wahlström, Adrian Wills, and Thomas B. Schön. Linearly Constrained Gaussian Processes. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Yoshifumi Kimura and Jackson R. Herring. Energy spectra of stably stratified turbulence. *Journal of Fluid Mechanics*, 698:19–50, 2012. doi:10.1017/jfm.2011.546.

Michel A. Kervaire. Non-Parallelizability of the n-Sphere for $n > 7$. *Proceedings of the National Academy of Sciences of the United States of America*, 44(3):280–283, 1958.

Byungsoo Kim, Vinicius C. Azevedo, Nils Thuerey, Theodore Kim, Markus Gross, and Barbara Solenthaler. Deep Fluids: A Generative Network for Parameterized Fluid Simulations. *Computer Graphics Forum*, 38(2):59–70, 2019.

Christopher Leon and Alexander Scheinker. Physics-constrained machine learning for electrodynamics without gauge ambiguity based on Fourier transformed Maxwell's equations. *Scientific Reports*, 14:14809, 2024. doi:10.1038/s41598-024-65650-9.

Xigui Li, Hongwei Zhang, Ruoxi Jiang, Deshu Chen, Chensen Lin, Limei Han, Yuan Qi, Xin Guo, and Yuan Cheng. Project and Generate: Divergence-Free Neural Operators for Incompressible Flows. arXiv:2603.24500, 2026.

Ning Liu, Yiming Fan, Xianyi Zeng, Milan Klöwer, Lu Zhang, and Yue Yu. Harnessing the Power of Neural Operators with Automatically Encoded Conservation Laws. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pages 30965–30997, 2024.

Arvind T. Mohan, Nicholas Lubbers, Misha Chertkov, and Daniel Livescu. Embedding Hard Physical Constraints in Neural Network Coarse-Graining of 3D Turbulence. *Physical Review Fluids*, 8:014604, 2023.

Xingyu Ni, Jingrui Xing, Xingqiao Li, Bin Wang, and Baoquan Chen. Representing Flow Fields with Divergence-Free Kernels for Reconstruction. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 8(4), Article 52, 2025.

Houman Owhadi. Gaussian Process Hydrodynamics. *Applied Mathematics and Mechanics (English Edition)*, 44(7):1175–1198, 2023.

Jack Richter-Powell, Yaron Lipman, and Ricky T. Q. Chen. Neural Conservation Laws: A Divergence-Free Perspective. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Varun Shankar, Gavin D. Portwood, Arvind T. Mohan, Peetak P. Mitra, Dilip Krishnamurthy, Christopher Rackauckas, Lucas A. Wilson, David P. Schmidt, and Venkatasubramanian Viswanathan. Validation and Parameterization of a Novel Physics-Constrained Neural Dynamics Model Applied to Turbulent Fluid Flow. *Physics of Fluids*, 34:115110, 2022.

Ramansh Sharma, Matthew Lowery, Houman Owhadi, and Varun Shankar. Fluids You Can Trust: Property-Preserving Operator Learning for Incompressible Flows. arXiv:2602.15472, 2026.

Jonathan Tompson, Kristofer Schlachter, Pablo Sprechmann, and Ken Perlin. Accelerating Eulerian Fluid Simulation With Convolutional Networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 3424–3433, 2017.

Gábor Tóth. The $\nabla \cdot B = 0$ constraint in shock-capturing magnetohydrodynamics codes. *Journal of Computational Physics*, 161(2):605–652, 2000.

Dima Tretiak, Arvind T. Mohan, and Daniel Livescu. Physics-Constrained Generative Adversarial Networks for 3D Turbulence. arXiv:2212.00217, 2022.

Fabian Waleffe. The nature of triad interactions in homogeneous turbulence. *Physics of Fluids A*, 4(2):350–363, 1992. doi:10.1063/1.858309.

Nils Wandel, Michael Weinmann, and Reinhard Klein. Learning Incompressible Fluid Dynamics from Scratch. In *International Conference on Learning Representations*, 2021.

Cheng Yang, Xubo Yang, and Xiangyun Xiao. Data-Driven Projection Method in Fluid Simulation. *Computer Animation and Virtual Worlds*, 27(3–4):415–424, 2016.

A Detailed Derivation of the Transverse Fourier Constraint

Let $u : \Omega \subset \mathbb{R}^D \to \mathbb{R}^D$ and define

$$\widehat{u}(k) = \int_\Omega u(x) e^{-ik\cdot x}\, dx. \tag{12}$$

We derive the constraint on $\widehat{u}(k)$ step by step. First, applying the product rule to the vector field $u(x)e^{-ik\cdot x}$ gives

$$\nabla\cdot\big(ue^{-ik\cdot x}\big) = e^{-ik\cdot x}\nabla\cdot u + u\cdot\nabla e^{-ik\cdot x}. \tag{13}$$

Since

$$\nabla e^{-ik\cdot x} = -ik\, e^{-ik\cdot x}, \tag{14}$$

Eq. (13) becomes

$$\nabla\cdot\big(ue^{-ik\cdot x}\big) = (\nabla\cdot u)e^{-ik\cdot x} - i(k^\top u)e^{-ik\cdot x}. \tag{15}$$

Integrating over Ω yields

$$\int_\Omega \nabla\cdot\big(ue^{-ik\cdot x}\big)\, dx = \int_\Omega (\nabla\cdot u)e^{-ik\cdot x}\, dx - i\int_\Omega (k^\top u)e^{-ik\cdot x}\, dx. \tag{16}$$

By the divergence theorem, the left-hand side is

$$\int_\Omega \nabla\cdot\big(ue^{-ik\cdot x}\big)\, dx = \int_{\partial\Omega} (u\cdot n)e^{-ik\cdot x}\, dS. \tag{17}$$

Because k is independent of x,

$$\int_\Omega (k^\top u)e^{-ik\cdot x}\, dx = k^\top \int_\Omega u(x)e^{-ik\cdot x}\, dx = k^\top \widehat{u}(k). \tag{18}$$

Substituting Eqs. (17) and (18) into Eq. (16) gives

$$\int_{\partial\Omega} (u\cdot n)e^{-ik\cdot x}\, dS = \int_\Omega (\nabla\cdot u)e^{-ik\cdot x}\, dx - ik^\top \widehat{u}(k), \tag{19}$$

and therefore

$$\boxed{k^\top \widehat{u}(k) = i\int_{\partial\Omega} (u\cdot n)e^{-ik\cdot x}\, dS - i\int_\Omega (\nabla\cdot u)e^{-ik\cdot x}\, dx.} \tag{20}$$

If the field is divergence-free,

$$\nabla\cdot u = 0 \quad \text{in } \Omega, \tag{21}$$

Eq. (20) reduces to

$$k^\top \widehat{u}(k) = i\int_{\partial\Omega} (u\cdot n)e^{-ik\cdot x}\, dS. \tag{22}$$

Hence the second sufficient condition is the vanishing weighted boundary flux

$$\int_{\partial\Omega} (u\cdot n)e^{-ik\cdot x}\, dS = 0. \tag{23}$$

For an impermeable boundary, $u\cdot n = 0$ pointwise on $\partial\Omega$, so Eq. (23) holds immediately. For a periodic box, opposite faces have opposite normals, while both u and the periodic Fourier factor $e^{-ik\cdot x}$ match across each paired face; their contributions therefore cancel exactly. Under either boundary condition,

$$\boxed{k^\top \widehat{u}(k) = 0.} \tag{24}$$

Thus, for every $k \neq 0$,

$$\widehat{u}(k) \in k^\perp := \{v \in \mathbb{C}^D : k^\top v = 0\}, \tag{25}$$

and since $k^\perp$ is a codimension-one subspace,

$$\boxed{\dim k^\perp = D - 1.} \tag{26}$$

Therefore each nonzero Fourier mode of a divergence-free D-component vector field satisfying the above boundary condition has exactly $D-1$ independent degrees of freedom.

B Operator Structure, Kernels, Invertibility, and Preserved Properties

This section makes the three realizations in the main text explicit. All Fourier-space dimensions below refer to the nonzero modes; the single zero mode is stored separately as $\bar{u} = \widehat{u}(0) \in \mathbb{R}^D$. Let $N_\times$ be the number of nonzero Fourier indices on the physical grid and $\widetilde{N}_\times$ the corresponding number on the extended grid.

B.1 Full statement of Theorem 1

Let $\Omega \subset \widetilde{\Omega}$. The field on Ω need not be periodic. Assume that a real divergence-free field u admits a periodic divergence-free continuation on $\widetilde{\Omega}$, and define the nonzero-mode decoder

$$A_E := M_D \widetilde{F}_D^{-1} \widetilde{E} \widetilde{F}_{D-1}. \tag{27}$$

Let

$$\mathcal{U}_E := \mathrm{Range}(A_E), \qquad \mathcal{C}_E := (\ker A_E)^\perp = \mathrm{Range}(A_E^*). \tag{28}$$

For every $u \in \mathcal{U}_E$, define the canonical forward transform (encoder) by

$$v_E^\star := \arg\min_v \frac{1}{2}\|v\|_2^2 \qquad \text{s.t.} \qquad A_E v = u, \tag{29}$$

with the closed form

$$v_E^\star = \mathcal{E}_E(u) = A_E^\dagger u \in \mathcal{C}_E. \tag{30}$$

The inverse transform (decoder) is

$$\mathcal{D}_E(v) := A_E v, \qquad v \in \mathcal{C}_E. \tag{31}$$

These maps satisfy

$$A_E A_E^\dagger u = u, \qquad u \in \mathcal{U}_E, \tag{32}$$

$$A_E^\dagger A_E v = v, \qquad v \in \mathcal{C}_E, \tag{33}$$

and therefore

$$\boxed{\mathcal{U}_E \underset{\mathcal{D}_E}{\overset{\mathcal{E}_E}{\rightleftarrows}} \mathcal{C}_E \subset L^2(\widetilde{\Omega}; \mathbb{R}^{D-1})} \tag{34}$$

is a one-to-one reversible representation.

The zero Fourier mode is not constrained by incompressibility and is retained separately as

$$\bar{u} := \widehat{u}(0) \in \mathbb{R}^D. \tag{35}$$

Hence the complete transform is

$$\boxed{u \longleftrightarrow (\bar{u}, v_E^\star)}. \tag{36}$$

If $\widehat{u}(0) = 0$, no additional zero-mode channel is required.

B.2 Per-mode and global operators

For every $k \neq 0$, let

$$E(k) \in \mathbb{R}^{D \times (D-1)}, \qquad E(k)^* E(k) = I_{D-1}, \qquad k^\top E(k) = 0. \tag{37}$$

The columns of $E(k)$ form an orthonormal basis of $k^\perp$. A full Householder reflection $Q_H(k) \in \mathbb{R}^{D \times D}$ may be written as

$$Q_H(k) = I_D - 2\frac{r_k r_k^\top}{r_k^\top r_k}, \qquad r_k = e_D - \frac{\kappa(k)}{|\kappa(k)|}, \tag{38}$$

with the paired convention $\kappa(-k) = \kappa(k)$; deleting the column aligned with $\kappa(k)$ gives $E(k)$. Hence $E(-k) = E(k)$ and Hermitian symmetry is preserved.

Stacking the mode-wise bases gives the rectangular block-diagonal operator

$$E_\times = \mathrm{blkdiag}_{k\neq 0}\, E(k) \in \mathbb{C}^{DN_\times \times (D-1)N_\times}. \tag{39}$$

It satisfies

$$E_\times^* E_\times = I_{(D-1)N_\times}, \qquad E_\times E_\times^* = P_\perp, \tag{40}$$

where $P_\perp = \mathrm{blkdiag}_{k\neq 0} P_k$ and $P_k = I_D - kk^\top/|k|^2$. Thus $E_\times E_\times^*$ is the identity on the divergence-free Fourier subspace.

Let $F_{D,N} = F_N \otimes I_D$ and $F_{D-1,N} = F_N \otimes I_{D-1}$ denote the component-wise unitary discrete Fourier transforms; on the nonzero-mode subspaces we use the same symbols with the zero coefficient omitted. Their dimensions are

$$F_{D,N} : \mathbb{C}^{DN} \to \mathbb{C}^{DN}, \qquad F_{D-1,N} : \mathbb{C}^{(D-1)N} \to \mathbb{C}^{(D-1)N}. \tag{41}$$

The paired symmetry $E(-k) = E(k)$ ensures that a real input field maps to Hermitian-symmetric reduced coefficients and therefore to a real $(D-1)$-component field after inverse FFT.

B.3 Case I: direct periodic transform

On the nonzero-mode subspaces the decoder and encoder are

$$A_0 = F_{D,N}^{-1} E_\times F_{D-1,N}, \qquad T_0 = A_0^{-1} = A_0^* = F_{D-1,N}^{-1} E_\times^* F_{D,N}. \tag{42}$$

Because the Fourier transforms are unitary and $E_\times^* E_\times = I$,

$$A_0^* A_0 = F_{D-1,N}^* E_\times^* E_\times F_{D-1,N} = I, \tag{43}$$

$$A_0 A_0^* = F_{D,N}^{-1} P_\perp F_{D,N}. \tag{44}$$

The second identity equals the identity when restricted to the physical divergence-free subspace. Therefore

$$\ker A_0 = \{0\}, \qquad A_0 : \mathcal{V}_{D-1} \xrightarrow{1:1} \mathcal{U}_{D,\mathrm{div}} \tag{45}$$

is a bijection. There is no optimization problem because $A_0 v = u$ has exactly one solution.

For any v_1, v_2,

$$\langle A_0 v_1, A_0 v_2 \rangle = \langle v_1, v_2 \rangle, \tag{46}$$

so norms, distances, and angles are preserved. This makes Case I particularly suitable as a representation or embedding for learning: with unitary normalization, the kinetic-energy norm is unchanged and, for any two samples,

$$\|A_0 v_1 - A_0 v_2\|_2^2 = \|v_1 - v_2\|_2^2, \tag{47}$$

so a squared-error discrepancy is identical before and after the transform. The encoder–decoder pair (T_0, A_0) is therefore a fixed analytic bijection from D divergence-free components to $D-1$ intrinsic components, rather than a latent embedding learned by a neural network; it introduces neither reconstruction error nor representation ambiguity on the admissible subspaces. Since spatial differentiation is the scalar Fourier multiplier ik_j on each mode,

$$(ik_j I_D) E(k) = E(k)(ik_j I_{D-1}), \tag{48}$$

and hence

$$\partial_{x_j} A_0 = A_0 \partial_{x_j}, \qquad \partial^\alpha A_0 = A_0 \partial^\alpha, \qquad \Delta A_0 = A_0 \Delta. \tag{49}$$

Consequently, for every Sobolev order s for which the norms are defined,

$$\|A_0 v\|_{H^s} = \|v\|_{H^s}, \qquad \|\partial^\alpha A_0 v\|_{L^2} = \|\partial^\alpha v\|_{L^2}. \tag{50}$$

Because A_0 is time independent,

$$\partial_t (A_0 v) = A_0 \partial_t v, \qquad \|\partial_t (A_0 v)\|_{L^2} = \|\partial_t v\|_{L^2}. \tag{51}$$

The same transform gives an exact coordinate change for dynamics. We first distinguish the two directions of the Case-I representation explicitly. The forward reduction (encoder) maps the physical divergence-free field to intrinsic coordinates,

$$\boxed{v = T_0 u = A_0^{-1} u = A_0^* u}, \tag{52}$$

whereas the inverse transform (decoder) reconstructs the physical field,

$$\boxed{u = A_0 v = T_0^{-1} v}. \tag{53}$$

Suppose the original D-component dynamics is

$$\boxed{\dot{u} = G_D(u)}, \qquad G_D : \mathcal{U}_{\text{div}} \to \mathcal{U}_{\text{div}}. \tag{54}$$

Differentiating the forward transform and substituting the inverse transform gives

$$\dot{v} = T_0 \dot{u} = T_0 G_D(u) = T_0 G_D(T_0^{-1} v), \tag{55}$$

so the exact reduced dynamics is

$$\boxed{\dot{v} = G_{D-1}(v)}, \qquad \boxed{G_{D-1} = T_0 \circ G_D \circ T_0^{-1} = A_0^{-1} \circ G_D \circ A_0}, \tag{56}$$

or, using $A_0^{-1} = A_0^*$,

$$\boxed{G_{D-1}(v) = A_0^* G_D(A_0 v)}. \tag{57}$$

Thus learning in the original representation means learning $G_D : u \mapsto \dot{u}$ on the D-component divergence-free state space, whereas our reduced formulation learns the conjugate operator $G_{D-1} : v \mapsto \dot{v}$ directly on the $(D-1)$-component intrinsic state space. The same conjugacy holds in discrete time: if $u_{n+1} = \Phi(u_n)$, then

$$v_{n+1} = T_0 \Phi(A_0 v_n). \tag{58}$$

Writing $G_{D-1}(v) = T_0 G_D(A_0 v)$, the continuous-time Jacobian obeys

$$J_{G_{D-1}}(v) = T_0 J_{G_D}(u) A_0. \tag{59}$$

Since $T_0 = A_0^*$ is unitary on the admissible spaces, this is unitary equivalence. Therefore

$$\operatorname{eig} J_{G_{D-1}} = \operatorname{eig} J_{G_D}, \qquad \sigma_i(J_{G_{D-1}}) = \sigma_i(J_{G_D}), \qquad \kappa(T_0) = 1. \tag{60}$$

Moreover,

$$\operatorname{tr} J_g = \operatorname{tr} J_f, \tag{61}$$

so phase-space divergence, and hence a Liouville-type divergence-free identity when present in the original coordinates, is preserved.

B.4 Case II: Fourier extension and canonicalization

Let $\widetilde{\Omega}$ contain $\widetilde{N} > N$ spatial points. The component-wise physical restriction is

$$M_D = M_x \otimes I_D \in \mathbb{R}^{DN \times D\widetilde{N}}, \tag{62}$$

where M_x selects the original N physical locations from the extended grid. Let

$$\widetilde{E}_\times \in \mathbb{C}^{D\widetilde{N}_\times \times (D-1)\widetilde{N}_\times} \tag{63}$$

be the extended block-diagonal transverse basis. The decoder is

$$A_E = M_D \widetilde{F}_{D,\widetilde{N}}^{-1} \widetilde{E}_\times \widetilde{F}_{D-1,\widetilde{N}}. \tag{64}$$

The Fourier/Householder part preceding M_D is injective, but M_D discards the extension region. Therefore nonzero reduced perturbations can decode to extended fields that vanish on Ω, giving

$$\ker A_E = \{z : M_D \widetilde{F}_D^{-1} \widetilde{E}_\times \widetilde{F}_{D-1} z = 0\}. \tag{65}$$

By rank-nullity,

$$\dim \ker A_E = (D-1)\widetilde{N}_\times - \operatorname{rank}(A_E). \tag{66}$$

If A_E is surjective onto an admissible physical state space of intrinsic dimension m, then $\operatorname{rank}(A_E) = m$ and the nullity is $(D-1)\widetilde{N}_\times - m$. Thus a single physical field has the complete inverse family

$$A_E^{-1}(u) = v_0 + \ker A_E, \tag{67}$$

which is one-to-many as an inverse relation, while the forward map is many-to-one.

The minimum-energy condition removes precisely this ambiguity:

$$v_E^\star = \arg\min_v \frac{1}{2}\|v\|_2^2 \quad \text{s.t.} \quad A_E v = u. \tag{68}$$

Its Moore–Penrose solution is

$$v_E^\star = A_E^\dagger u = A_E^*(A_E A_E^*)^{-1} u \tag{69}$$

when the indicated inverse exists on the admissible range, and

$$v_E^\star \in \mathrm{Range}(A_E^*) = (\ker A_E)^\perp. \tag{70}$$

Hence

$$A_E : (\ker A_E)^\perp \xrightarrow{1:1} \mathrm{Range}(A_E) \tag{71}$$

is bijective. Numerically, if $A_E = U\Sigma V^*$, the thresholded pseudoinverse uses $1/\sigma_i$ only for $\sigma_i > \tau\sigma_{\max}$ and sets the remaining inverse singular values to zero.

Because restriction is not unitary,

$$A_E^* A_E = \widetilde{F}_{D-1}^* \widetilde{E}_\times^* \widetilde{F}_D M_D^* M_D \widetilde{F}_D^{-1} \widetilde{E}_\times \widetilde{F}_{D-1} \neq I \tag{72}$$

in general. Thus canonical bijectivity does not imply standard L^2 isometry or angle preservation. The forward decoder still commutes with spatial derivatives in the interior of Ω, but the minimum-norm encoder $A_E^\dagger$ need not commute with spatial differentiation. Since $A_E^\dagger$ is fixed in time, however,

$$\partial_t(A_E^\dagger u) = A_E^\dagger \partial_t u. \tag{73}$$

On the canonical subspace the reduced dynamics remains an exact coordinate representation,

$$\dot{v} = A_E^\dagger f(A_E v), \tag{74}$$

so Jacobian eigenvalues are preserved by similarity, while Euclidean singular values and conditioning generally are not.

B.5 Case III: extension with matched spectral dimension

Let K be a conjugate-symmetric retained set with $|K| = N_\times$ nonzero modes. The coefficient injection

$$J_K \in \mathbb{C}^{(D-1)\widetilde{N}_\times \times (D-1)N_\times} \tag{75}$$

places the retained reduced coefficients at their extended-grid Fourier indices; $P_K^F = J_K J_K^*$ is the corresponding spectral projector. The decoder is

$$A_K = M_D \widetilde{F}_D^{-1} \widetilde{E}_\times J_K F_{D-1,N}. \tag{76}$$

Its reduced domain has the same intrinsic nonzero-mode dimension $(D-1)N_\times$ as the target divergence-free state space. Dimension matching alone does not prove invertibility; Case III therefore chooses/accepts a mode set K only when

$$\mathrm{rank}(A_K) = (D-1)N_\times, \qquad \mathrm{Range}(A_K) = \mathcal{U}_{D,\mathrm{adm}}. \tag{77}$$

Under this full-rank condition,

$$\ker A_K = \{0\}, \qquad v_K = A_K^{-1} u, \tag{78}$$

so no minimum-norm optimization is needed. The spectral selection J_K commutes with spatial Fourier multipliers on the retained modes, and the fixed transform commutes with time differentiation. However, because M_D and spectral truncation remain present, A_K is not generally unitary and does not automatically preserve the full-space L^2/H^s geometry.

B.6 Summary of the three realizations

The three cases differ only in how the admissible reduced coordinates are related to the physical field:

$$\begin{array}{lll} \text{Case I:} & \ker A_0 = \{0\}, & v = A_0^{-1} u, \\ \text{Case II:} & \ker A_E \neq \{0\} \text{ in general}, & v = A_E^\dagger u \in (\ker A_E)^\perp, \\ \text{Case III:} & \ker A_K = \{0\} \text{ under the full-rank condition}, & v = A_K^{-1} u. \end{array} \tag{79}$$

Thus Case I is a unitary coordinate change, Case II is a redundant extension followed by canonicalization, and Case III removes the extension overparameterization by spectral dimension matching plus a full-rank requirement. In all three cases, once restricted to the stated invertible spaces, the physical dynamics can be transported without information loss to $(D-1)$-component coordinates.

C Global Smoothness of the Transverse Frame and Empirical Effect of Gauge Seams

The minimal representation requires more than identifying the transverse subspace $k^\perp$: it also requires choosing concrete coordinates inside that subspace. For every nonzero Fourier mode, let

$$n(k) = \frac{k}{|k|} \in S^{D-1}, \qquad E(k) = \big[e_1(k), \dots, e_{D-1}(k)\big], \tag{80}$$

where the columns of $E(k)$ are an orthonormal basis of $k^\perp$. Thus the construction attaches a $(D-1)$-dimensional coordinate frame to every direction on the wavevector sphere S^{D-1}. Ideally, nearby wavevector directions would receive nearby frames: if $n(k_i)$ and $n(k_j)$ are close on the sphere, then $E(k_i)$ and $E(k_j)$ would also be close.

C.1 Parallelizability and the intrinsic global obstruction

A smooth choice of a complete tangent frame over the whole sphere is exactly the classical problem of whether the sphere is *parallelizable*. Apart from the trivial zero-dimensional case, the only parallelizable spheres are

$$\boxed{S^1, \qquad S^3, \qquad S^7,} \tag{81}$$

as established by the classical sphere-parallelizability results (Bott & Milnor, 1958; Kervaire, 1958). Since our wavevector directions lie on S^{D-1}, a globally smooth complete real transverse frame can therefore exist only for

$$\boxed{D = 2, \qquad D = 4, \qquad D = 8} \tag{82}$$

for the nontrivial dimensions considered here. In the common three-dimensional case, $D = 3$, the direction sphere is S^2. The hairy-ball obstruction is already sufficient to rule out even one globally nonvanishing continuous tangent vector field on S^2, and hence a globally smooth two-vector transverse frame is impossible. Consequently, in three dimensions the loss of global smoothness is intrinsic to any single global real frame construction; replacing Householder by another single global frame can move or reshape the problematic set, but cannot remove the obstruction.

The paired reality convention used in this paper imposes the additional requirement $E(-k) = E(k)$ through a canonical representative of each pair $\{k, -k\}$. This convention guarantees real physical-space reduced channels, but the canonical half-space selection can introduce an additional coordinate seam. Therefore the smoothness of a particular implementation can be more restrictive than sphere parallelizability alone.

C.2 How the obstruction appears in the Householder construction

The Householder frame used in Appendix B fixes a reference direction e_D and sets, for the canonical unit direction $n = \kappa/|\kappa|$,

$$Q_H(n) = I_D - 2\frac{(e_D - n)(e_D - n)^\top}{\|e_D - n\|_2^2}, \tag{83}$$

with a separately defined value at the aligned direction $n = e_D$. The transverse frame $E(n)$ is obtained from the first $D-1$ columns of $Q_H(n)$. The aligned direction is a concrete location where the coordinate choice becomes singular. To see why assigning a special value at that single point does not make the surrounding frame continuous, approach it as

$$n(\theta, \omega) = \cos\theta\, e_D + \sin\theta\, \omega, \qquad \omega \perp e_D, \qquad \|\omega\|_2 = 1. \tag{84}$$

As $\theta \to 0$, the normalized Householder vector approaches $-\omega$, so the limiting reflection contains the factor

$$I_D - 2\omega\omega^\top, \tag{85}$$

which depends on the tangential approach direction ω. Thus two wavevector directions can become arbitrarily close to one another while their selected transverse coordinate frames remain separated by a finite rotation or sign change.

We quantify this purely geometric effect by the angular frame roughness

$$\theta_{ij} = \arccos\big(\mathrm{clip}\big(n_i^\top n_j, -1, 1\big)\big)\,, \tag{86}$$

$$\boxed{R_E(i,j) = \frac{\|E(k_i) - E(k_j)\|_F}{\theta_{ij}}.} \tag{87}$$

Equation (87) measures how much the chosen transverse coordinate axes change per unit angular displacement in Fourier space. In our direct geometric diagnostic, the squared neighbor-to-neighbor frame difference near the reference-axis singular region was approximately 1.98, compared with approximately 0.023 in ordinary regions, a ratio of about 85. This confirms that the selected frame can vary far more rapidly near the seam even though the physical transverse subspace $k^\perp$ itself changes smoothly.

C.3 Why frame roughness can create rough reduced Fourier coordinates

The previous subsection concerns the coordinate frame $E(k)$ itself. The object learned by the reduced model is instead

$$\widehat{a}(k) = E(k)^\top \widehat{u}(k). \tag{88}$$

These are distinct questions. A rapidly varying frame does not automatically imply that $\widehat{a}(k)$ is rough for every physical field, but it makes such roughness possible. In particular, even when two neighboring physical Fourier coefficients satisfy

$$\widehat{u}(k_i) \approx \widehat{u}(k_j), \tag{89}$$

a large change in $E(k)$ can still produce a large change between $\widehat{a}(k_i)$ and $\widehat{a}(k_j)$. In elementary terms, the physical vector can change only slightly while the coordinate axes used to measure it rotate rapidly.

To isolate this mechanism, we first use a controlled Fourier field whose directional dependence is smooth by construction. For three deterministic coordinate probes $c_m = e_m$, $m = 1, 2, 3$, define

$$P_\perp(k) = I - \frac{kk^\top}{|k|^2}, \qquad h(|k|) = \exp\left[-\frac{(|k| - 6)^2}{8}\right], \tag{90}$$

$$\boxed{\widehat{u}_m(k) = h(|k|)P_\perp(k)c_m, \qquad \widehat{u}_m(0) = 0.} \tag{91}$$

Because $k^\top P_\perp(k) = 0$, every probe is exactly divergence-free. Away from $k = 0$, both the radial envelope and the projector vary smoothly with the wavevector direction, so Eq. (91) provides a clean control in which any additional sharp variation after encoding can be attributed to the coordinate frame rather than to native turbulent fluctuations.

For the same neighboring-mode pairs used in the geometric diagnostic, define the normalized physical and reduced differences

$$D_u(i,j) = \frac{\|\widehat{u}(k_i) - \widehat{u}(k_j)\|_2}{\sqrt{\left(\|\widehat{u}(k_i)\|_2^2 + \|\widehat{u}(k_j)\|_2^2\right)/2} + 10^{-12}}, \tag{92}$$

$$D_a(i,j) = \frac{\|\widehat{a}(k_i) - \widehat{a}(k_j)\|_2}{\sqrt{\left(\|\widehat{a}(k_i)\|_2^2 + \|\widehat{a}(k_j)\|_2^2\right)/2} + 10^{-12}}. \tag{93}$$

The seam set is fixed *before* looking at D_u or D_a: we take the top 1% of neighboring pairs ranked by R_E as seam pairs and the lowest 50% as normal pairs. We then summarize the excess variation by

$$S_u = \frac{\mathrm{median}(D_u \mid \mathrm{seam})}{\mathrm{median}(D_u \mid \mathrm{normal})}, \qquad S_a = \frac{\mathrm{median}(D_a \mid \mathrm{seam})}{\mathrm{median}(D_a \mid \mathrm{normal})}, \tag{94}$$

and report

$$\boxed{\Gamma = \frac{S_a}{S_u}.} \tag{95}$$

Here S_u asks how much rougher the original Fourier field is near the geometric seam than in ordinary regions, while S_a asks the same question after the transverse coordinate transform. Thus $\Gamma > 1$ is direct evidence that the coordinate choice adds seam-localized variation beyond that already present in the physical Fourier field. The three smooth probes expose this effect clearly, confirming that the mechanism is real when the underlying Fourier field is locally smooth.

C.4 Real forced-HIT data: native spectral variability dominates the added seam effect

We next test whether the same mechanism is appreciable on actual turbulent data. We use the publicly released coarsened filtered-DNS data accompanying the Implicit Adaptive Fourier Neural Operator (IAFNO) study (Jiang et al., 2026). The benchmark is *forced homogeneous isotropic turbulence* (forced HIT) in a periodic three-dimensional box. The public forced-HIT file used in our experiments, `FGR1_nofilter_vel_50g_600p_gap200_LES_32.npy`, contains 50 trajectories, 600 saved snapshots per trajectory, a 32^3 spatial grid, and three velocity components.

For the frame-smoothness diagnostic we use five held-out trajectories and sample twenty temporally separated snapshots from each trajectory, giving

$$\boxed{5 \times 20 = 100 \text{ snapshots.}} \tag{96}$$

For every snapshot we compute $\widehat{u}(k)$, encode $\widehat{a}(k) = E(k)^\top \widehat{u}(k)$, and evaluate Eqs. (92)–(95) on exactly the same preselected seam and normal pairs. The seam set is never re-selected from the turbulent data. Sampling across five held-out trajectories prevents the conclusion from depending on a single turbulent realization, and the twenty separated times per trajectory prevent it from depending on a single instant. We additionally use 1000 bootstrap resamples of the snapshot-level statistic to estimate a 95% confidence interval.

The resulting real-data amplification ratio is

$$\boxed{\Gamma_{\mathrm{HIT}} = \frac{S_a^{\mathrm{HIT}}}{S_u^{\mathrm{HIT}}} \approx 0.9957,} \tag{97}$$

with a 95% bootstrap interval of approximately

$$\boxed{[0.9903,\ 1.0051].} \tag{98}$$

The interval contains 1 and the point estimate is itself essentially unity. Therefore, for these forced-HIT snapshots, we do not observe a systematic increase in seam-localized Fourier roughness after encoding.

The reason is visible in the denominator of this comparison. In the synthetic probes, neighboring physical Fourier coefficients are deliberately made locally smooth, so a rapid change of coordinate frame stands out strongly. In an individual turbulent snapshot, neighboring Fourier coefficients already exhibit substantial mode-to-mode variability in amplitude, phase, and transverse direction. The additional variation introduced by the Householder seam is therefore small relative to the native spectral variability. The empirical conclusion is consequently more specific than the topological statement: the global frame obstruction is genuine and can be exposed on a smooth control field, while its additional roughness is weak on the tested real forced-HIT distribution.

C.5 Scope of the limitation

The seam discussed above is a loss of smoothness with respect to the *wavevector direction*; it should not be interpreted as a literal discontinuity of the finite-resolution reduced field in physical space. At fixed numerical resolution, each reduced channel is a finite Fourier series and is therefore smooth in x. Moreover, the modewise transform remains orthonormal on $k^\perp$, so the direct periodic representation remains exactly invertible and preserves the represented modal energy. The practical concern is instead that the coordinate description can make neighboring Fourier coefficients vary more rapidly, can complicate the representation of spatial rotations, and can produce a less localized physical-space transform kernel. Our controlled experiment confirms that this coordinate effect exists, whereas the forced-HIT experiment shows that its additional Fourier roughness is negligible relative to the natural mode-to-mode variability of the tested turbulent data.

C.6 Globally smooth direct transverse frames in $D = 2, 4, 8$

The preceding nullspace calculations are algebraic, while a separate geometric issue arises when a minimal representation chooses an explicit orthonormal basis of $k^\perp$ for every nonzero wavevector.

Writing $n = k/|k| \in S^{D-1}$, such a choice is a global orthonormal tangent frame on S^{D-1}. The classical sphere-parallelizability result implies that, among the nontrivial cases, a globally smooth complete frame exists precisely for

$$\boxed{D = 2, \qquad D = 4, \qquad D = 8,} \tag{99}$$

because S^1, S^3, and S^7 are parallelizable (Bott & Milnor, 1958; Kervaire, 1958). In contrast, for $D = 3$ the direction sphere is S^2, which is not parallelizable. Hence no construction can choose two everywhere continuous, everywhere nonvanishing, linearly independent real vectors spanning $k^\perp$ over all directions. A Householder frame can place the problematic set at a convenient reference direction, but in three dimensions the existence of at least one singular direction, seam, or discontinuous transition is intrinsic and cannot be removed by changing the reference vector or by replacing Householder with another single global real frame.

For $D = 2$, a global frame is immediate. For $k = (k_1, k_2)^\top \neq 0$, define

$$\boxed{E_2(k) = \frac{1}{|k|}\begin{bmatrix} -k_2 \\ k_1 \end{bmatrix}.} \tag{100}$$

Then $k^\top E_2(k) = 0$ and $E_2(k)^\top E_2(k) = 1$. Equation (100) is smooth for every $k \neq 0$ and therefore gives a single globally smooth transverse coordinate on S^1, with no distinguished pole.

For $D = 4$, the same idea has a particularly simple quaternionic realization. Identify

$$q(k) = k_1 + k_2\mathbf{i} + k_3\mathbf{j} + k_4\mathbf{k} \in \mathbb{H}, \tag{101}$$

where $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1$ and $\mathbf{ij} = \mathbf{k}$, $\mathbf{jk} = \mathbf{i}$, $\mathbf{ki} = \mathbf{j}$. Right multiplication by the three imaginary quaternion units produces the three tangent directions $q\mathbf{i}$, $q\mathbf{j}$, and $q\mathbf{k}$. In real coordinates this gives

$$\boxed{E_4(k) = \frac{1}{|k|}\begin{bmatrix} -k_2 & -k_3 & -k_4 \\ k_1 & -k_4 & k_3 \\ k_4 & k_1 & -k_2 \\ -k_3 & k_2 & k_1 \end{bmatrix}.} \tag{102}$$

A direct calculation gives

$$k^\top E_4(k) = 0, \qquad E_4(k)^\top E_4(k) = I_3. \tag{103}$$

Equivalently, multiplication by a unit quaternion is an orthogonal map, so it carries the orthonormal set $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ to an orthonormal basis of the tangent space at $q/|q|$. Every entry of Eq. (102) is linear in k divided only by $|k|$; on S^3 the denominator is identically one. Thus this quaternion construction is globally smooth on the entire direction sphere, with no Householder reference-direction singularity, polar singularity, or chart seam. In four dimensions, Householder is therefore optional rather than necessary: the quaternion frame provides a simpler global construction that does not introduce the discontinuity associated with a fixed Householder reference direction.

For $D = 8$, the corresponding global construction is obtained from the octonions. Write an octonion as a Cayley–Dickson pair of quaternions,

$$\mathbb{O} = \mathbb{H} \oplus \mathbb{H}\ell, \qquad (a, b)(c, d) = \big(ac - \overline{d}\, b,\ da + b\overline{c}\big), \tag{104}$$

and use the basis

$$\begin{aligned} 1, \quad e_1 = (\mathbf{i}, 0), \quad e_2 = (\mathbf{j}, 0), \quad e_3 = (\mathbf{k}, 0), \quad e_4 = (0, 1), \\ e_5 = (0, \mathbf{i}), \quad e_6 = (0, \mathbf{j}), \quad e_7 = (0, \mathbf{k}). \end{aligned} \tag{105}$$

For

$$q(k) = k_1 + k_2e_1 + k_3e_2 + k_4e_3 + k_5e_4 + k_6e_5 + k_7e_6 + k_8e_7, \tag{106}$$

take the seven columns to be the real coordinate vectors of $qe_1, \ldots, qe_7$. With the above multiplication convention this yields

$$\boxed{E_8(k) = \frac{1}{|k|}\begin{bmatrix} -k_2 & -k_3 & -k_4 & -k_5 & -k_6 & -k_7 & -k_8 \\ k_1 & -k_4 & k_3 & -k_6 & k_5 & k_8 & -k_7 \\ k_4 & k_1 & -k_2 & -k_7 & -k_8 & k_5 & k_6 \\ -k_3 & k_2 & k_1 & -k_8 & k_7 & -k_6 & k_5 \\ k_6 & k_7 & k_8 & k_1 & -k_2 & -k_3 & -k_4 \\ -k_5 & k_8 & -k_7 & k_2 & k_1 & k_4 & -k_3 \\ -k_8 & -k_5 & k_6 & k_3 & -k_4 & k_1 & k_2 \\ k_7 & -k_6 & -k_5 & k_4 & k_3 & -k_2 & k_1 \end{bmatrix}.} \tag{107}$$

The octonion norm is multiplicative. Therefore, for unit q, multiplication by q preserves Euclidean inner products, and the seven vectors $qe_1, \ldots, qe_7$ are mutually orthonormal and orthogonal to q. Consequently,

$$k^\top E_8(k) = 0, \qquad E_8(k)^\top E_8(k) = I_7. \tag{108}$$

As in the quaternion case, Eq. (107) is a single algebraic formula on all of $\mathbb{R}^8 \setminus \{0\}$ and hence a globally smooth frame on S^7. It has no distinguished reference axis, no pole, and no Householder-type singularity or chart transition. Different standard sign conventions for the octonion multiplication merely permute or sign-flip the seven columns and give equivalent global frames.

Thus the three special dimensions admit explicit direct constructions,

$$D = 2 : \mathbb{C}, \qquad D = 4 : \mathbb{H}, \qquad D = 8 : \mathbb{O}, \tag{109}$$

that span $k^\perp$ globally and smoothly. In $D = 4$ and $D = 8$, there is consequently no need to force the transverse coordinates through a Householder construction: the quaternionic and octonionic frames avoid the artificial singular direction introduced by that particular parameterization. The impossibility in $D = 3$ is qualitatively different because it is topological rather than a defect of Householder. These smoothness statements concern the frame on the oriented direction sphere itself; if an implementation additionally imposes an antipodal pairing convention such as $E(-k) = E(k)$ to obtain real reduced channels, that extra gauge convention is a separate constraint and can introduce its own seam even though the quaternionic or octonionic frame itself is globally smooth.

D Fourier-Space Derivation and Explicit Redundancy of the NCL Parameterizations

This appendix starts from the two constructions in Neural Conservation Laws (Richter-Powell et al., 2022) and derives their Fourier-space form step by step. NCL uses the dimension symbol n; we write D to match the notation of this paper. The NCL formulas below are their constructions, while the Fourier-space nullspace analysis is our derivation.

D.1 NCL matrix-field construction in physical space

NCL begins with differential forms. For an arbitrary $(D-2)$-form μ, the identity $d^2 = 0$ gives

$$0 = d^2\mu = d(d\mu), \tag{110}$$

and NCL defines

$$v = \star d\mu, \tag{111}$$

which is divergence-free. Writing

$$\mu = \frac{1}{2} \sum_{i,j=1}^{D} \mu_{ij} \star (dx_i \wedge dx_j), \qquad \mu_{ji} = -\mu_{ij}, \tag{112}$$

gives the coordinate expression

$$v_i(x) = \sum_{j=1}^{D} \frac{\partial \mu_{ij}(x)}{\partial x_j}. \tag{113}$$

NCL then sets $A_{ij} = \mu_{ij}$, so $A^\top = -A$, and obtains its first practical construction,

$$\boxed{v_i(x) = \sum_{j=1}^{D} \partial_j A_{ij}(x)} \qquad \Longleftrightarrow \qquad v = \mathrm{div}_{\mathrm{row}}\, A. \tag{114}$$

The divergence vanishes directly because mixed derivatives commute and $A_{ji} = -A_{ij}$:

$$\nabla \cdot v = \sum_{i,j} \partial_i \partial_j A_{ij} = \frac{1}{2} \sum_{i,j} \partial_i \partial_j \left(A_{ij} + A_{ji}\right) = 0. \tag{115}$$

Thus the matrix field has $D(D-1)/2$ independent scalar components, but its decoded field is constrained to the divergence-free subspace.

D.2 FOURIER TRANSFORM OF THE MATRIX-FIELD CONSTRUCTION

Use the Fourier convention

$$\widehat{f}(k) = \int f(x) e^{-ik\cdot x}\, dx, \tag{116}$$

for which differentiation becomes multiplication by ik_j,

$$\widehat{\partial_j f}(k) = ik_j \widehat{f}(k). \tag{117}$$

Applying Eq. (117) to Eq. (114) gives, component by component,

$$\widehat{v}_i(k) = \sum_{j=1}^{D} ik_j \widehat{A}_{ij}(k) = i \sum_{j=1}^{D} \widehat{A}_{ij}(k) k_j. \tag{118}$$

Stacking all components yields the matrix-vector product

$$\boxed{\widehat{v}(k) = i\widehat{A}(k)k.} \tag{119}$$

Because $\widehat{A}^\top = -\widehat{A}$,

$$k^\top \widehat{v}(k) = i\, k^\top \widehat{A}(k) k = 0, \tag{120}$$

so every decoded mode lies in $k^\perp$. For each nonzero k, NCL's matrix decoder is therefore the linear map

$$T_k : \mathfrak{so}(D) \to k^\perp, \qquad T_k(\widehat{A}) = i\widehat{A}k. \tag{121}$$

D.3 WHERE THE MATRIX-FIELD REDUNDANCY IS LOCATED

The redundancy is exactly the kernel of T_k. Let $C(k)$ be any antisymmetric matrix satisfying

$$C(k)^\top = -C(k), \qquad C(k)k = 0. \tag{122}$$

Then adding C changes the representation but not the physical field:

$$T_k(\widehat{A} + C) = i(\widehat{A} + C)k = i\widehat{A}k + iCk = T_k(\widehat{A}). \tag{123}$$

Hence

$$\boxed{\ker T_k = \{C \in \mathfrak{so}(D) : Ck = 0\}.} \tag{124}$$

For a desired transverse mode $\widehat{v} \in k^\perp$, one explicit representation is

$$\widehat{A}_0(k) = -\frac{i}{|k|^2} \left(\widehat{v}(k) k^\top - k \widehat{v}(k)^\top \right), \tag{125}$$

because $i\widehat{A}_0 k = \widehat{v}$. Therefore the complete family of matrix representations of the *same* physical mode is

$$\boxed{\widehat{A}(k) = \widehat{A}_0(k) + C(k), \qquad C(k)^\top = -C(k), \quad C(k)k = 0.} \tag{126}$$

Equation (126) shows explicitly how redundancy can be added: $C(k)$ may vary arbitrarily inside this nullspace while $\widehat{v}(k)$ remains exactly unchanged.

To make this visible entry by entry, rotate coordinates so that $k = |k|e_D$. Every antisymmetric matrix has the block form

$$\widehat{A} = \begin{pmatrix} B & a \\ -a^\top & 0 \end{pmatrix}, \qquad B^\top = -B, \qquad a \in \mathbb{C}^{D-1}. \tag{127}$$

Multiplication by k gives

$$\widehat{A}k = |k| \begin{pmatrix} a \\ 0 \end{pmatrix}, \qquad \widehat{v} = i|k| \begin{pmatrix} a \\ 0 \end{pmatrix}. \tag{128}$$

Thus only a affects the physical field; the entire block $B \in \mathfrak{so}(D-1)$ is invisible. Equivalently, one may add

$$C = \begin{pmatrix} \Delta B & 0 \\ 0 & 0 \end{pmatrix}, \qquad \Delta B^\top = -\Delta B, \tag{129}$$

with arbitrary ΔB, and the output does not change because $Ck = 0$.

The rank is $D - 1$ and therefore, by rank-nullity,

$$\boxed{\dim \ker T_k = \frac{D(D-1)}{2} - (D-1) = \frac{(D-1)(D-2)}{2}.} \tag{130}$$

The matrix representation is minimal only in $D = 2$; the invisible dimensions are 1 in $D = 3$, 3 in $D = 4$, and grow quadratically with D.

Explicit 3D example. For $k = \kappa e_3$, write

$$\widehat{A} = \begin{pmatrix} 0 & c & a_1 \\ -c & 0 & a_2 \\ -a_1 & -a_2 & 0 \end{pmatrix}. \tag{131}$$

Then

$$\widehat{v} = i\widehat{A}k = i\kappa \begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix}, \tag{132}$$

which is completely independent of c. Thus c can be replaced by $c + \delta c$ for any δc without changing the physical Fourier mode. In this coordinate system the redundant addition is explicitly

$$C = \begin{pmatrix} 0 & \delta c & 0 \\ -\delta c & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad Ck = 0. \tag{133}$$

D.4 NCL VECTOR-FIELD CONSTRUCTION IN PHYSICAL SPACE

NCL's second construction starts from an arbitrary vector field $b : \mathbb{R}^D \to \mathbb{R}^D$. In differential-form notation NCL sets $\mu = \delta\nu$ and obtains

$$v = \star d\delta\nu. \tag{134}$$

Writing ν in the $(D-1)$-form basis gives, up to the sign convention of the differential-form operators,

$$\delta\nu = \frac{1}{2} \sum_{i,j=1}^{D} \left(\frac{\partial \nu_i}{\partial x_j} - \frac{\partial \nu_j}{\partial x_i} \right) \star (dx_i \wedge dx_j). \tag{135}$$

NCL therefore gives the practical matrix formula

$$\boxed{A = J_b - J_b^\top,} \tag{136}$$

where $(J_b)_{ij} = \partial_j b_i$. Equation (136) by itself only defines the antisymmetric matrix field A; the physical vector field is obtained only after applying the row-wise divergence decoder

$$v_i = \sum_{j=1}^{D} \partial_j A_{ij}. \tag{137}$$

Substituting the entries of Eq. (136),

$$A_{ij} = \partial_j b_i - \partial_i b_j, \tag{138}$$

into Eq. (137) gives, component by component,

$$v_i = \sum_j \partial_j \left(\partial_j b_i - \partial_i b_j \right) \tag{139}$$

$$= \sum_j \partial_j^2 b_i - \partial_i \sum_j \partial_j b_j \tag{140}$$

$$= \Delta b_i - \partial_i (\nabla \cdot b). \tag{141}$$

Thus Eq. (142) below is not an independent assumption and is not equivalent to Eq. (136) alone; it is exactly the decoded vector field obtained from Eq. (136) together with the row-wise divergence in Eq. (137):

$$\boxed{v = \Delta b - \nabla(\nabla \cdot b) = \left(\Delta I - \nabla\nabla^\top\right) b.} \tag{142}$$

Here I is the $D \times D$ identity matrix and the operator $\nabla\nabla^\top$ acts on a vector field as

$$(\nabla\nabla^\top)b = \nabla(\nabla \cdot b). \tag{143}$$

Before transforming the matrix formula, it is useful to Fourier-transform Eq. (142) directly. With the convention $\widehat{\partial_j f}(k) = ik_j \widehat{f}(k)$, where the lowercase $i = \sqrt{-1}$ is the imaginary unit,

$$\widehat{\Delta b}(k) = -|k|^2 \widehat{b}(k), \qquad \widehat{\nabla(\nabla \cdot b)}(k) = -kk^\top \widehat{b}(k). \tag{144}$$

Therefore

$$\widehat{v}(k) = \left(-|k|^2 I + kk^\top\right) \widehat{b}(k) \tag{145}$$

$$= -|k|^2 \left(I - \frac{kk^\top}{|k|^2}\right) \widehat{b}(k) \tag{146}$$

$$= \boxed{-|k|^2 P_k \widehat{b}(k)}, \qquad P_k = I - \frac{kk^\top}{|k|^2}. \tag{147}$$

The matrix P_k is exactly the orthogonal projector onto $k^\perp$: it satisfies

$$P_k^\top = P_k, \qquad P_k^2 = P_k, \qquad P_k k = 0, \tag{148}$$

and for every z with $k^\top z = 0$ one has $P_k z = z$. Hence the projector in Eq. (147) is not introduced by analogy; it follows algebraically from Fourier-transforming the physical-space operator in Eq. (142).

D.5 Fourier transform of the vector-field construction

The same Fourier formula can be derived independently from the matrix representation in Eq. (136). From the derivative rule, the Fourier transform of the Jacobian is

$$\widehat{J}_b(k) = i\widehat{b}(k)k^\top. \tag{149}$$

Hence Eq. (136) becomes

$$\widehat{A}(k) = i\widehat{b}(k)k^\top - ik\widehat{b}(k)^\top \tag{150}$$

$$= i\left(\widehat{b}(k)k^\top - k\widehat{b}(k)^\top\right). \tag{151}$$

Substituting this into the matrix decoder $\widehat{v} = i\widehat{A}k$ gives

$$\widehat{v}(k) = i\, i\left(\widehat{b}k^\top - k\widehat{b}^\top\right) k \tag{152}$$

$$= -\left(|k|^2\widehat{b} - k(k^\top\widehat{b})\right) \tag{153}$$

$$= -\left(|k|^2 I - kk^\top\right)\widehat{b} \tag{154}$$

$$= \boxed{-|k|^2 P_k \widehat{b}(k)}, \qquad P_k = I - \frac{kk^\top}{|k|^2}. \tag{155}$$

Thus the NCL vector construction is exactly a mode-wise Leray projection of the arbitrary vector $\widehat{b}(k)$, followed by the scalar factor $-|k|^2$.

D.6 Where the vector-field redundancy is located

Decompose the arbitrary vector into transverse and longitudinal parts,

$$\widehat{b}(k) = \widehat{b}_\perp(k) + \alpha(k)k, \qquad k^\top\widehat{b}_\perp = 0. \tag{156}$$

Since $P_k k = 0$ and $P_k \widehat{b}_\perp = \widehat{b}_\perp$,

$$\widehat{v}(k) = -|k|^2 \widehat{b}_\perp(k). \tag{157}$$

The entire longitudinal component $\alpha(k)k$ is therefore invisible. For a prescribed physical mode $\widehat{v} \in k^\perp$, one particular representation is

$$\widehat{b}_0(k) = -\frac{\widehat{v}(k)}{|k|^2}, \tag{158}$$

and every representation of the same mode is

$$\boxed{\widehat{b}(k) = \widehat{b}_0(k) + \alpha(k)k, \qquad \alpha(k) \in \mathbb{C}.} \tag{159}$$

This is the explicit many-to-one direction: changing $\alpha(k)$ changes the NCL representation but leaves $\widehat{v}(k)$ exactly unchanged.

For $k = \kappa e_D$ this becomes particularly transparent:

$$\widehat{b} = (b_1, \ldots, b_{D-1}, b_D)^\top, \qquad \widehat{v} = -\kappa^2 (b_1, \ldots, b_{D-1}, 0)^\top. \tag{160}$$

The last component b_D can take any value without affecting the output. In three dimensions, for example,

$$(b_1, b_2, b_3) \quad \text{and} \quad (b_1, b_2, b_3 + \delta) \tag{161}$$

produce exactly the same $\widehat{v}$ for every δ.

D.7 Interpretation of the two nullspaces

The two NCL constructions reach the same transverse physical space but carry different redundant variables. The matrix construction maps

$$\mathfrak{so}(D) \ni \widehat{A} \longmapsto i\widehat{A}k \in k^\perp \tag{162}$$

with a $(D-1)(D-2)/2$-dimensional nullspace, while the vector construction maps

$$\mathbb{C}^D \ni \widehat{b} \longmapsto -|k|^2 P_k \widehat{b} \in k^\perp \tag{163}$$

with the one-dimensional nullspace $\mathrm{span}\{k\}$. The matrix construction is therefore not literally a projection of a vector; it is a surjective linear decoder from antisymmetric matrices with a large kernel. The vector construction is literally a scaled Leray projection. In both cases many distinct internal representations correspond to one identical physical Fourier mode, whereas a direct transverse parameterization uses only the $D-1$ coordinates of $k^\perp$.

Finally, at $k = 0$ all derivative factors vanish, so both derivative-based NCL constructions produce zero for that mode; this is the constant-mode exception also reflected in NCL's universality statement.

E DFK as a Finite-Basis Realization of the NCL Vector Construction

This appendix makes explicit the mathematical relation between the DFK construction (Ni et al., 2025) and the vector-field parameterization of Neural Conservation Laws (Richter-Powell et al., 2022) derived in Appendix D. The connection is established algebraically rather than by analogy. The DFK construction represents a free D-component vector field with a finite collection of localized radial kernels and then applies the second-order transverse operator $-I\Delta + \nabla\nabla^\top$. The NCL vector-$b$ construction represents a free D-component vector field with a neural parameterization and applies the opposite-sign operator $\Delta I - \nabla\nabla^\top$. Thus, after identifying the two free vector fields, the decoded divergence-free fields differ only by an overall sign. In three dimensions, the DFK operator also admits a double-curl rewrite; this is an algebraic consequence of the standard vector identity and is derived below. A single curl is already sufficient for the divergence-free guarantee.

E.1 DFK: KERNEL EXPANSION AND THE DOUBLE-CURL IDENTITY

The DFK construction (Ni et al., 2025) starts from a scalar radial kernel $\phi_i(x)$ and defines the matrix-valued divergence-free kernel

$$\Psi_i(x) = \left(-I\Delta + \nabla\nabla^\top\right)\phi_i(x). \tag{164}$$

With a constant vector coefficient $\omega_i \in \mathbb{R}^D$, one basis contribution is

$$u_i^{\mathrm{DFK}}(x) = \Psi_i(x)\omega_i = \left(-I\Delta + \nabla\nabla^\top\right)\left(\phi_i(x)\omega_i\right), \tag{165}$$

and the reconstructed field is

$$u_{\mathrm{DFK}}(x) = \sum_{i=1}^{M} u_i^{\mathrm{DFK}}(x). \tag{166}$$

We use two typographically similar but different symbols throughout this appendix. The uppercase I denotes the $D \times D$ identity matrix, whereas the lowercase $i = \sqrt{-1}$ that appears after Fourier transformation is the imaginary unit. Since the Laplacian is a scalar differential operator applied componentwise,

$$I\Delta = \Delta I, \tag{167}$$

so the notations $-I\Delta$ and $-\Delta I$ mean exactly the same operator. Likewise,

$$(\nabla\nabla^\top)F = \nabla(\nabla \cdot F). \tag{168}$$

With these conventions, the divergence of the DFK operator is

$$\nabla \cdot \left[\left(-\Delta I + \nabla\nabla^\top\right)F\right] = -\Delta(\nabla \cdot F) + \nabla \cdot \nabla(\nabla \cdot F) \tag{169}$$

$$= -\Delta(\nabla \cdot F) + \Delta(\nabla \cdot F) = 0. \tag{170}$$

Thus every basis contribution and every linear combination are divergence-free.

In three dimensions, the same second-order operator has a double-curl representation. Componentwise,

$$\left[\nabla \times (\nabla \times F)\right]_i = \partial_i\left(\sum_j \partial_j F_j\right) - \sum_j \partial_j^2 F_i \tag{171}$$

$$= \left[\nabla(\nabla \cdot F) - \Delta F\right]_i. \tag{172}$$

Therefore

$$\boxed{\nabla \times (\nabla \times F) = \nabla(\nabla \cdot F) - \Delta F = \left(-\Delta I + \nabla\nabla^\top\right)F.} \tag{173}$$

Equation (170) and Eq. (173) are related but are not the same statement: Eq. (173) identifies the full operator in three dimensions, while Eq. (170) states only that its output has zero divergence. Equation (173) immediately implies the divergence-free property because the divergence of any curl vanishes.

Setting $F = \phi_i\omega_i$ gives

$$\boxed{u_i^{\mathrm{DFK}} = \nabla \times \nabla \times (\phi_i\omega_i)}. \tag{174}$$

Thus the "double-curl" description is simply an algebraic rewrite of the second-order DFK operator in three dimensions; the DFK itself is introduced through Eq. (164).

The Fourier-space form follows directly from the same identity. Under $\widehat{\nabla \times F}(k) = ik \times \widehat{F}(k)$, each spatial derivative contributes one factor of the lowercase imaginary unit i. Hence

$$\widehat{\nabla \times \nabla \times F}(k) = (ik) \times \left[(ik) \times \widehat{F}(k)\right] \tag{175}$$

$$= -k \times \left(k \times \widehat{F}(k)\right) \tag{176}$$

$$= -\left[k\left(k^\top \widehat{F}(k)\right) - |k|^2 \widehat{F}(k)\right] \tag{177}$$

$$= \left(|k|^2 I - kk^\top\right)\widehat{F}(k) \tag{178}$$

$$= \boxed{|k|^2 P_k \widehat{F}(k)}, \qquad P_k = I - \frac{kk^\top}{|k|^2}. \tag{179}$$

The penultimate equality is the vector triple-product identity $k\times(k\times a) = k(k^\top a) - |k|^2 a$. The final equality is just factorization of $|k|^2$; P_k is the orthogonal projector onto $k^\perp$ because it annihilates the component parallel to k and leaves every vector orthogonal to k unchanged. Thus the Fourier-space projector is exactly the transform of the physical-space operator $-\Delta I + \nabla\nabla^\top$, not a separate modeling assumption.

This identity also makes clear why two curls are not required merely to enforce incompressibility. The same work separately compares the one-curl Curl Kernel (Ni et al., 2025) of the form

$$u_{\mathrm{curl}} = \nabla \times (\phi\omega), \tag{180}$$

which already satisfies

$$\boxed{\nabla \cdot u_{\mathrm{curl}} = \nabla \cdot \big(\nabla \times (\phi\omega)\big) = 0.} \tag{181}$$

Therefore the second curl in Eq. (174) is not an additional divergence-free requirement. It changes the basis produced from the same underlying scalar kernel, including its local geometry and spectral weighting.

Now define the free vector field represented by the kernels,

$$b_{\mathrm{DFK}}(x) = \sum_{i=1}^{M} \phi_i(x)\omega_i. \tag{182}$$

Linearity of the differential operator gives

$$u_{\mathrm{DFK}} = \sum_{i=1}^{M} \left(-I\Delta + \nabla\nabla^\top\right)\left(\phi_i\omega_i\right) \tag{183}$$

$$= \left(-I\Delta + \nabla\nabla^\top\right) \sum_{i=1}^{M} \phi_i\omega_i \tag{184}$$

$$= \boxed{\left(-I\Delta + \nabla\nabla^\top\right) b_{\mathrm{DFK}}.} \tag{185}$$

In three dimensions, Eqs. (173) and (185) also give

$$u_{\mathrm{DFK}} = \nabla \times \nabla \times b_{\mathrm{DFK}}. \tag{186}$$

The many kernels therefore form a finite localized basis for an otherwise free D-component auxiliary vector field $b(x)$.

E.2 Step-by-step equivalence with the NCL vector-b construction

Appendix D derives NCL's vector construction from

$$A = J_b - J_b^\top, \tag{187}$$

followed by the row-wise divergence

$$(v_{\mathrm{NCL}})_i = \sum_{j=1}^{D} \partial_j A_{ij}. \tag{188}$$

Substituting $A_{ij} = \partial_j b_i - \partial_i b_j$ into Eq. (188) gives

$$(v_{\mathrm{NCL}})_i = \sum_j \partial_j \left(\partial_j b_i - \partial_i b_j\right) \tag{189}$$

$$= \sum_j \partial_j^2 b_i - \partial_i \sum_j \partial_j b_j \tag{190}$$

$$= \Delta b_i - \partial_i(\nabla \cdot b). \tag{191}$$

Hence

$$\boxed{v_{\mathrm{NCL}} = \left(\Delta I - \nabla\nabla^\top\right) b.} \tag{192}$$

Equation (192) is the physical-space NCL operator derived directly from Eqs. (187)–(188).

Compare Eq. (192) with the DFK result in Eq. (185):

$$\left(-I\Delta + \nabla\nabla^\top\right) = -\left(\Delta I - \nabla\nabla^\top\right). \tag{193}$$

Therefore, if the same free vector field is used, $b = b_{\mathrm{DFK}}$,

$$\boxed{u_{\mathrm{DFK}} = -v_{\mathrm{NCL}}.} \tag{194}$$

The overall sign can be absorbed into the kernel coefficients ω_i, so the two constructions generate exactly the same function class after the free vector field is identified.

The same equivalence is explicit in Fourier space. Using

$$\widehat{\Delta b}(k) = -|k|^2\widehat{b}(k), \qquad \widehat{\nabla(\nabla\cdot b)}(k) = -kk^\top\widehat{b}(k), \tag{195}$$

Eq. (192) becomes

$$\widehat{v}_{\mathrm{NCL}}(k) = \left(-|k|^2 I + kk^\top\right)\widehat{b}(k) \tag{196}$$

$$= -\left(|k|^2 I - kk^\top\right)\widehat{b}(k) \tag{197}$$

$$= \boxed{-|k|^2 P_k \widehat{b}(k)}, \qquad P_k = I - \frac{kk^\top}{|k|^2}. \tag{198}$$

For the DFK operator,

$$\widehat{u}_{\mathrm{DFK}}(k) = \left(|k|^2 I - kk^\top\right)\widehat{b}_{\mathrm{DFK}}(k) \tag{199}$$

$$= \boxed{|k|^2 P_k \widehat{b}_{\mathrm{DFK}}(k)}. \tag{200}$$

Equations (198) and (200) are exact negatives of one another. They have the same range $k^\perp$ and the same null direction $\mathrm{span}\{k\}$. This is the precise sense in which the DFK construction is a special kernel-basis realization of the NCL vector-b construction.

E.3 Relation to the deterministic *Project* branch

The same mode-wise redundancy appears in the deterministic *Project* branch (Li et al., 2026). That method first produces an unconstrained D-component Fourier mode $\widehat{q}(k) \in \mathbb{C}^D$ and then applies the Leray projector

$$\widehat{u}_{\mathrm{P\&G}}(k) = P_k\widehat{q}(k), \qquad P_k = I - \frac{kk^\top}{|k|^2}. \tag{201}$$

Writing $\widehat{q}(k) = \widehat{q}_\perp(k) + \alpha(k)k$ gives

$$P_k\widehat{q}(k) = \widehat{q}_\perp(k), \qquad P_k\big(\alpha(k)k\big) = 0. \tag{202}$$

Thus the deterministic *Project* branch (Li et al., 2026), the NCL vector-b construction (Richter-Powell et al., 2022), and the DFK construction (Ni et al., 2025) share the same mode-wise geometric pattern: they introduce a D-component ambient representation containing one longitudinal degree of freedom that is invisible to the final divergence-free field. Their operators differ in spectral scaling: the deterministic *Project* branch uses P_k, NCL uses $-|k|^2P_k$, and DFK uses $|k|^2P_k$. In all three cases the admissible range is $k^\perp$ and the longitudinal direction $\mathrm{span}\{k\}$ is removed. Our representation differs by never introducing this redundant longitudinal degree of freedom: it starts directly from $D-1$ coordinates in $k^\perp$.

E.4 What differs: the representation of the free vector field

NCL uses a coordinate neural network to parameterize the free D-component field,

$$b(x) = b_\theta(x) \in \mathbb{R}^D. \tag{203}$$

The DFK construction (Ni et al., 2025) instead uses the finite kernel expansion in Eq. (182),

$$b_{\mathrm{DFK}}(x) = \sum_{i=1}^{M} \phi_i(x)\omega_i, \tag{204}$$

with trainable kernel locations, scales, and vector weights. Thus both constructions first specify a free D-component vector field and then apply the same second-order transverse differential operator up to sign; the difference is the function class used to represent that free field.

E.5 KERNEL COUNTS AND TRAINABLE DEGREES OF FREEDOM

The finite-basis nature of the DFK representation is visible in the reported parameter counts (Ni et al., 2025). In the two-dimensional Kármán vortex-street experiment, DFK-Wen4 uses $5{,}367$ kernels and $26{,}835$ trainable parameters. This corresponds to five trainable quantities per kernel: two center coordinates, one scale, and a two-component vector weight. Their Curl Kernel comparison uses $6{,}794$ kernels and $27{,}176$ trainable parameters.

In the three-dimensional analytic-vortex experiment, the kernel methods use $21{,}117$ kernels and $147{,}819$ trainable parameters. The count is seven parameters per kernel,

$$(x_i, y_i, z_i, h_i, \omega_{ix}, \omega_{iy}, \omega_{iz}), \tag{205}$$

namely three center coordinates, one scale, and a three-component vector weight. These numbers make the interpretation concrete: DFK represents the auxiliary vector field b with a large collection of localized basis functions and then applies the transverse operator.

E.6 COMPARISON WITH THE MINIMAL TRANSVERSE REPRESENTATION

The NCL vector form and DFK retain a free D-component auxiliary vector field before the transverse operator is applied. Mode by mode, the longitudinal part of that auxiliary field is invisible because

$$P_k\left(\widehat{b}_\perp + \alpha k\right) = \widehat{b}_\perp. \tag{206}$$

The NCL matrix form instead retains an antisymmetric matrix field $A(x) \in \mathfrak{so}(D)$ and applies its row-wise divergence. Mode by mode, this is the surjective linear decoder $\widehat{A}(k) \mapsto i\widehat{A}(k)k \in k^\perp$, rather than a vector projection. Our representation instead stores a real $(D-1)$-component reduced field $a(x,t)$. At each time t, a spatial Fourier transform produces its mode coefficients $\widehat{a}(k,t)$, these $D-1$ coefficients are mapped by the transverse basis $E(k)$ to the physical Fourier coefficient $\widehat{u}(k,t)$, and an inverse spatial Fourier transform returns the real D-component field. The comparison is therefore

$$\begin{aligned}
\text{NCL vector form} \quad &: \quad b_\theta(x) \in \mathbb{R}^D \xrightarrow{\Delta I - \nabla\nabla^\top} u_{\text{NCL-vec}}(x) \in \mathbb{R}^D \quad , \\
&\quad \text{equivalently:} \quad \widehat{b}_\theta(k) \in \mathbb{C}^D \xrightarrow{-|k|^2 P_k} \widehat{u}_{\text{NCL-vec}}(k) \in k^\perp \\
\text{DFK} \quad &: \quad b_{\text{DFK}}(x) = \sum_i \phi_i(x)\omega_i \in \mathbb{R}^D \xrightarrow{-\Delta I + \nabla\nabla^\top} u_{\text{DFK}}(x) \in \mathbb{R}^D, \\
&\quad \text{equivalently:} \quad \widehat{b}_{\text{DFK}}(k) \in \mathbb{C}^D \xrightarrow{|k|^2 P_k} \widehat{u}_{\text{DFK}}(k) \in k^\perp \\
\text{NCL matrix form} \quad &: \quad A_\theta(x) \in \mathfrak{so}(D) \xrightarrow{\text{div}_{\text{row}}} u_{\text{NCL-mat}}(x) \in \mathbb{R}^D \quad , \\
&\quad \text{equivalently:} \quad \widehat{A}_\theta(k) \in \mathfrak{so}(D,\mathbb{C}) \xrightarrow{i\widehat{A}_\theta(k)k} \widehat{u}_{\text{NCL-mat}}(k) \in k^\perp \\
\text{ours} \quad &: \quad a(x,t) \in \mathbb{R}^{D-1} \xrightarrow{\mathcal{F}_x} \widehat{a}(k,t) \in \mathbb{C}^{D-1} \\
&\qquad\qquad \xrightarrow{E(k)} \widehat{u}(k,t) \in k^\perp \xrightarrow{\mathcal{F}_x^{-1}} u(x,t) \in \mathbb{R}^D.
\end{aligned} \tag{207}$$

Here $\mathfrak{so}(D) = \{A \in \mathbb{R}^{D\times D} : A^\top = -A\}$ and $\mathfrak{so}(D,\mathbb{C})$ is its complex Fourier-space counterpart; $P_k = I - kk^\top/|k|^2$. Also, $\mathcal{F}_x$ denotes the spatial Fourier transform only, so time is not Fourier transformed. Thus $a(x,t)$ is the real reduced field seen by the learning model, while $\widehat{a}(k,t)$ and $\widehat{u}(k,t)$ are internal Fourier-space representations used by the fixed analytic transform. The paired $k/-k$ convention preserves Hermitian symmetry, so the final inverse transform is real. The NCL vector and DFK rows use the same transverse second-order construction up to an overall sign and differ mainly in how the free D-component vector field is represented. The NCL matrix row is instead a redundant antisymmetric-matrix decoder with a large kernel. The final row removes these auxiliary null directions and learns directly with only $D-1$ real channels.

F Navier–Stokes and Magnetohydrodynamics in Minimal Transverse Coordinates

This appendix makes explicit how the divergence-free state representation developed in this paper changes the governing equations themselves. The main point is structural: an incompressible evolution is usually written as a vector evolution equation together with a separate divergence constraint. If the state is parameterized directly in the transverse Fourier subspace, the divergence constraint is satisfied identically and therefore no longer appears as an independent equation. In addition, pressure gradients lie entirely in the longitudinal Fourier direction and vanish after transformation to the reduced coordinates. We first show this for the Navier–Stokes equations and then for incompressible magnetohydrodynamics (MHD).

F.1 Incompressible Navier–Stokes in D dimensions

For a constant-density incompressible fluid, the velocity field $u(x,t) \in \mathbb{R}^D$ satisfies

$$\partial_t u + (u\cdot\nabla)u = -\nabla p + \nu\Delta u + f, \tag{208}$$

$$\nabla\cdot u = 0, \tag{209}$$

where the pressure has been divided by the constant density, ν is the kinematic viscosity, and f is a body force. Thus the standard incompressible formulation consists of a D-component evolution equation coupled to the scalar constraint in Eq. (209). The pressure acts as the Lagrange multiplier that keeps the velocity evolution inside the divergence-free subspace (Chorin, 1968).

F.1.1 Two dimensions: scalar vorticity and stream function

In two spatial dimensions, introduce the constant 90° rotation matrix

$$R = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad R^\top R = I, \qquad R^\top = -R, \tag{210}$$

and define a scalar stream function ψ by

$$u = R\nabla\psi \equiv \nabla^\perp\psi. \tag{211}$$

Because the Hessian of ψ is symmetric while R is antisymmetric,

$$\nabla\cdot u = \operatorname{tr}\big(R\nabla^2\psi\big) = 0 \tag{212}$$

identically. Hence incompressibility is built into the stream-function representation rather than imposed as a separate equation.

Let the scalar vorticity be $\omega = \operatorname{curl}_{2D} u$. With the convention in Eq. (211),

$$\omega = \Delta\psi. \tag{213}$$

Taking the curl of Eq. (208) eliminates the pressure gradient and gives

$$\partial_t\omega + (u\cdot\nabla)\omega = \nu\Delta\omega + \operatorname{curl}_{2D} f. \tag{214}$$

Substituting $u = R\nabla\psi$ and $\omega = \Delta\psi$ produces a single scalar evolution equation,

$$\boxed{\partial_t\Delta\psi + \big(R\nabla\psi\cdot\nabla\big)\Delta\psi = \nu\Delta^2\psi + \operatorname{curl}_{2D} f.} \tag{215}$$

The pair "momentum equation + incompressibility constraint" has therefore been replaced by one scalar stream-function/vorticity equation, with incompressibility satisfied automatically.

For the inviscid unforced limit $\nu = 0$, Eq. (214) reduces to material transport of scalar vorticity,

$$\partial_t\omega + (u\cdot\nabla)\omega = 0. \tag{216}$$

Under periodic boundaries, sufficient decay at infinity, or boundary conditions that remove the corresponding flux terms, this implies conservation of both kinetic energy and enstrophy,

$$\mathcal{E} = \frac{1}{2}\int |u|^2\,dx, \qquad \mathcal{Z} = \frac{1}{2}\int \omega^2\,dx. \tag{217}$$

Here the second invariant is *enstrophy*, not entropy. For viscous Navier–Stokes, these quantities are dissipative rather than exactly conserved; in particular, in the unforced periodic setting,

$$\frac{d\mathcal{E}}{dt} = -\nu\int |\nabla u|^2\,dx, \qquad \frac{d\mathcal{Z}}{dt} = -\nu\int |\nabla\omega|^2\,dx. \tag{218}$$

F.1.2 THREE DIMENSIONS: VORTEX STRETCHING

In three dimensions, the vorticity is the vector

$$\omega = \nabla \times u. \tag{219}$$

Taking the curl of the incompressible momentum equation yields

$$\boxed{\partial_t \omega + (u\cdot\nabla)\omega = (\omega\cdot\nabla)u + \nu\Delta\omega + \nabla \times f.} \tag{220}$$

The additional nonlinear term

$$(\omega\cdot\nabla)u \tag{221}$$

is the vortex-stretching term. It is absent from the two-dimensional scalar-vorticity equation. Consequently, even in the inviscid unforced limit, the three-dimensional enstrophy is not generally conserved:

$$\frac{d}{dt}\frac{1}{2}\int |\omega|^2\,dx = \int \omega\cdot(\omega\cdot\nabla)u\,dx, \tag{222}$$

whereas kinetic energy remains conserved for smooth inviscid solutions under the same no-flux assumptions. Three-dimensional Euler also conserves helicity $\int u\cdot\omega\,dx$ under the standard corresponding assumptions. With viscosity, kinetic energy is dissipated and Eq. (222) acquires the additional negative viscous contribution $-\nu\int|\nabla\omega|^2dx$.

The dimensional distinction can therefore be summarized without expanding component equations:

	2D incompressible flow	3D incompressible flow
vorticity	scalar	vector
vortex stretching	absent	$(\omega\cdot\nabla)u$
inviscid energy	conserved	conserved
inviscid enstrophy	conserved	not generally conserved

(223)

F.2 THE $D \to D-1$ TRANSVERSE FORMULATION OF NAVIER–STOKES

The stream-function formulation above is a special two-dimensional way to build incompressibility into the state variable. The transverse representation in this paper provides an analogous construction in arbitrary spatial dimension D.

For each nonzero Fourier mode k, let

$$E(k) \in \mathbb{R}^{D\times(D-1)}, \qquad E(k)^*E(k) = I_{D-1}, \qquad E(k)E(k)^* = P_k = I - \frac{kk^\top}{|k|^2}, \tag{224}$$

so the columns of $E(k)$ form an orthonormal basis of $k^\perp$. Using the notation of the main text, define

$$A_0 = \mathcal{F}_D^{-1}E\mathcal{F}_{D-1}, \qquad A_0^* = \mathcal{F}_{D-1}^{-1}E^*\mathcal{F}_D. \tag{225}$$

Write the physical velocity as

$$u = A_0 a, \qquad a = A_0^* u, \qquad a(x,t) \in \mathbb{R}^{D-1}. \tag{226}$$

For every nonzero mode,

$$k^\top \widehat{u}(k) = k^\top E(k)\widehat{a}(k) = 0, \tag{227}$$

so $\nabla\cdot u = 0$ is an identity of the representation.

Now apply A_0^* to Eq. (208). Because $E(k)^*k = 0$,

$$A_0^*\nabla p = 0. \tag{228}$$

Because A_0 and the Laplacian are both Fourier multipliers and $E(k)^*E(k) = I_{D-1}$,

$$A_0^*\Delta A_0 a = \Delta a. \tag{229}$$

Therefore the incompressible Navier–Stokes system reduces to the single $(D-1)$-component evolution equation

$$\boxed{\partial_t a + A_0^*[((A_0 a)\cdot\nabla)(A_0 a)] = \nu\Delta a + A_0^* f.} \tag{230}$$

No separate equation $\nabla\cdot u = 0$ is required, and the pressure variable also disappears from the reduced evolution because its gradient is purely longitudinal. The physical velocity is recovered analytically as $u = A_0 a$. Thus, modulo the separately handled zero Fourier mode described in the main text, the usual pair

$$\begin{cases} D\text{-component momentum equation}, \\ \nabla\cdot u = 0 \end{cases} \tag{231}$$

is represented directly by one evolution equation for a $(D-1)$-component state.

F.3 From Maxwell plus fluid dynamics to incompressible MHD

We next make the same structure explicit for MHD. The physical derivation is naturally three-dimensional because the cross product is three-dimensional; after eliminating the cross products, the final advective/tension form extends directly to the D-dimensional transverse notation.

F.3.1 Electromagnetic part

In the nonrelativistic, low-frequency MHD approximation, the Maxwell equations needed for magnetic evolution are

$$\nabla\cdot B = 0, \tag{232}$$

$$\partial_t B = -\nabla\times E, \tag{233}$$

$$\nabla\times B = \mu_0 J, \tag{234}$$

where the displacement-current term is neglected in Ampère's law. A resistive Ohm law for a moving conducting fluid is

$$E + u\times B = \sigma^{-1} J, \tag{235}$$

with conductivity σ. The ideal-MHD limit is $\sigma\to\infty$, giving $E + u\times B = 0$.

Substituting Eq. (235) into Faraday's law and then using Eq. (234) gives

$$\partial_t B = \nabla\times(u\times B) - \sigma^{-1}\nabla\times J \tag{236}$$

$$= \nabla\times(u\times B) + \eta_m \Delta B, \tag{237}$$

where

$$\eta_m = \frac{1}{\mu_0\sigma} \tag{238}$$

is the magnetic diffusivity and $\nabla\cdot B = 0$ has been used. Taking the divergence of Eq. (237) shows that an initially divergence-free magnetic field remains divergence-free; nevertheless, numerical schemes often require explicit control of this constraint (Tóth, 2000).

F.3.2 Fluid part and Lorentz force

Mass conservation is

$$\partial_t\rho + \nabla\cdot(\rho u) = 0. \tag{239}$$

For constant density, this becomes

$$\nabla\cdot u = 0. \tag{240}$$

The fluid momentum equation with the electromagnetic force is

$$\rho\left[\partial_t u + (u\cdot\nabla)u\right] = -\nabla p + \rho\nu\Delta u + J\times B + \rho f, \tag{241}$$

where the electric force density is neglected in the usual quasi-neutral nonrelativistic MHD limit. Ampère's law and $\nabla\cdot B = 0$ give the vector identity

$$J\times B = \frac{1}{\mu_0}(\nabla\times B)\times B = \frac{1}{\mu_0}\left[(B\cdot\nabla)B - \nabla\frac{|B|^2}{2}\right]. \tag{242}$$

Thus the Lorentz force separates into magnetic tension and a magnetic-pressure gradient.

Define the Alfvén-scaled magnetic field and total pressure by

$$b = \frac{B}{\sqrt{\mu_0\rho}}, \qquad \Pi = \frac{p}{\rho} + \frac{|b|^2}{2}, \tag{243}$$

for constant ρ. Using

$$\nabla \times (u \times b) = (b\cdot\nabla)u - (u\cdot\nabla)b + u(\nabla\cdot b) - b(\nabla\cdot u), \tag{244}$$

and imposing both divergence constraints, incompressible resistive MHD becomes

$$\partial_t u + (u\cdot\nabla)u = -\nabla\Pi + (b\cdot\nabla)b + \nu\Delta u + f, \tag{245}$$

$$\partial_t b + (u\cdot\nabla)b = (b\cdot\nabla)u + \eta_m\Delta b, \tag{246}$$

$$\nabla\cdot u = 0, \tag{247}$$

$$\nabla\cdot b = 0. \tag{248}$$

This is the four-equation structure relevant to the present representation: two vector evolution equations plus two divergence constraints. The two constraints have different physical origins: $\nabla\cdot u = 0$ follows from mass conservation at constant density, whereas $\nabla\cdot B = 0$ is Gauss's law for magnetism.

F.4 Reducing incompressible MHD from four equations to two

Because both u and b are divergence-free vector fields on the same domain, they can be encoded independently by the same transverse frame. Let

$$u = A_0 a, \qquad b = A_0 c, \qquad a, c \in \mathbb{R}^{D-1}. \tag{249}$$

Then

$$\nabla\cdot(A_0 a) = 0, \qquad \nabla\cdot(A_0 c) = 0 \tag{250}$$

identically for all nonzero modes. Applying A_0^* to Eqs. (245)–(246), and again using $A_0^*\nabla\Pi = 0$, gives

$$\boxed{\partial_t a + A_0^*[(u\cdot\nabla)u - (b\cdot\nabla)b] = \nu\Delta a + A_0^* f,} \tag{251}$$

$$\boxed{\partial_t c + A_0^*[(u\cdot\nabla)b - (b\cdot\nabla)u] = \eta_m\Delta c,} \tag{252}$$

with the analytic reconstructions

$$u = A_0 a, \qquad b = A_0 c. \tag{253}$$

Therefore the standard four-equation incompressible MHD structure

$$\begin{cases} \text{velocity evolution,} \\ \text{magnetic-field evolution,} \\ \nabla\cdot u = 0, \\ \nabla\cdot b = 0 \end{cases} \tag{254}$$

is represented by only two reduced vector evolution equations, one for a and one for c. Both divergence constraints are satisfied by construction, and the total-pressure gradient is removed automatically by the transverse coordinate map. In an operator-learning implementation, a neural model may therefore evolve (a, c) directly in $2(D-1)$ real channels and decode them analytically to the physical (u, b) fields, rather than predicting $2D$ ambient channels and subsequently enforcing two separate solenoidal constraints.

G Direct Random Generation of Divergence-Free Fields in Transverse Coordinates

The same $D \to D-1$ representation can be used to generate random divergence-free fields directly, without first generating an unconstrained D-component field and then projecting it. For every nonzero Fourier mode k, let

$$E(k) \in \mathbb{R}^{D\times(D-1)}, \qquad E(k)^* E(k) = I_{D-1}, \qquad k^\top E(k) = 0, \tag{255}$$

where $E(k)$ is the transverse orthonormal frame obtained from the Householder construction used in the main method.

Choose one representative from each Fourier pair $\{k, -k\}$. In the intrinsic $(D-1)$-dimensional coordinates, sample independent real and imaginary Gaussian vectors,

$$\xi_R(k), \xi_I(k) \sim \mathcal{N}(0, I_{D-1}), \qquad \widehat{a}(k) = S(|k|)\big(\xi_R(k) + i\xi_I(k)\big), \tag{256}$$

where $S(|k|)$ controls the spectral decay. For example,

$$S(|k|) = \left(1 + \frac{|k|^2}{k_0^2}\right)^{-\alpha/2} \qquad \text{or} \qquad S(|k|) = \exp\left[-\left(\frac{|k|}{k_c}\right)^q\right]. \tag{257}$$

Larger α or q suppresses high frequencies more strongly, while smaller values retain more high-frequency energy.

Decode each sampled reduced coefficient directly into the physical D-component Fourier coefficient,

$$\widehat{u}(k) = E(k)\widehat{a}(k) \in \mathbb{C}^D. \tag{258}$$

No projection is required, because

$$k^\top \widehat{u}(k) = k^\top E(k)\widehat{a}(k) = 0. \tag{259}$$

To obtain a real physical-space field, impose Hermitian symmetry on the paired mode,

$$\widehat{u}(-k) = \overline{\widehat{u}(k)}, \tag{260}$$

with the zero mode chosen separately and self-conjugate Nyquist modes taken real when present. The final field is then

$$\boxed{u(x) = \mathcal{F}^{-1}\widehat{u}, \qquad \nabla\cdot u = 0} \tag{261}$$

up to numerical roundoff. Thus the complete generation pipeline is

$$\boxed{\begin{aligned} &\mathcal{N}(0, I_{D-1}) \xrightarrow{S(|k|)} \widehat{a}(k) \in \mathbb{C}^{D-1} \xrightarrow{E(k)} \widehat{u}(k) \in k^\perp \subset \mathbb{C}^D \\ &\widehat{u}(k) \xrightarrow{\text{Hermitian pairing}} \widehat{u}(-k) = \overline{\widehat{u}(k)} \\ &\widehat{u}(-k) \xrightarrow{\mathcal{F}^{-1}} u(x) \in \mathbb{R}^D, \quad \nabla\cdot u = 0. \end{aligned}} \tag{262}$$

The random field is therefore generated directly in the intrinsic transverse subspace, with exactly $D-1$ random degrees of freedom per nonzero Fourier mode and no post-hoc Leray projection.

H SPECTRAL MEANING OF THE 2D NAVIER–STOKES PREDICTION LOSSES

This appendix summarizes why the five 2D Navier–Stokes prediction variables used in our comparisons induce different spectral weightings. The equations below describe the squared-L^2 error numerator; common averaging or relative-L^2 normalization does not change the corresponding mode weighting within each representation.

Velocity–vorticity relation. For every nonzero Fourier mode of a two-dimensional incompressible velocity field,

$$\widehat{u}(k) = \frac{ik^{\perp}}{|k|^2}\widehat{\omega}(k), \qquad |\widehat{\omega}(k)|^2 = |k|^2|\widehat{u}(k)|^2. \tag{263}$$

Let $e_u(k) = \widehat{u}_{\mathrm{pred}}(k) - \widehat{u}_{\mathrm{true}}(k)$ and $e_\omega(k) = \widehat{\omega}_{\mathrm{pred}}(k) - \widehat{\omega}_{\mathrm{true}}(k)$. For divergence-free errors, Eq. (263) gives

$$|e_\omega(k)|^2 = |k|^2|e_u(k)|^2. \tag{264}$$

Five prediction variables. The corresponding losses can therefore be read as follows:

Model	Predicted variable	Equivalent spectral weighting
Original FNO-ω	ω	$L_\omega = \sum_k \lvert e_\omega\rvert^2 = \sum_k \lvert k\rvert^2\lvert e_u\rvert^2$
A (raw)	u	$L_A = \sum_k \lvert e_u\rvert^2$
B (Leray)	$P_k u_{\mathrm{raw}}$	$L_B = \sum_k \lvert P_k e_{\mathrm{raw}}\rvert^2$
C (potential)	ψ	$L_C = \sum_k \lvert e_\psi\rvert^2 = \sum_{k\neq 0} \frac{\lvert e_u\rvert^2}{\lvert k\rvert^2}$
$D2$ (transverse)	a	$L_{D2} = \sum_k \lvert e_a\rvert^2 = \sum_k \lvert e_u\rvert^2$

(265)

Here $P_k = I - kk^{\top}/|k|^2$. Model B predicts an ambient velocity and removes its longitudinal component by projection, so only the transverse part of the raw error reaches the projected output loss. For model C, $u = \nabla^{\perp}\psi$ implies $\widehat{u} = ik^{\perp}\widehat{\psi}$, hence $|e_\psi|^2 = |e_u|^2/|k|^2$. For D2, the orthonormal transverse coordinate is isometric to velocity, so $|e_a|^2 = |e_u|^2$.

Thus the main spectral distinction is immediate: direct vorticity prediction weights velocity error by $|k|^2$ and therefore emphasizes high wavenumbers; the potential representation weights it by $1/|k|^2$ and therefore emphasizes low wavenumbers; raw velocity and D2 use the ordinary velocity L^2 weighting. D2 differs from raw velocity in representation, not in the ideal L^2 norm: it stores only the single admissible transverse degree of freedom in 2D.

Radial-versus-phase ablation. In two dimensions the D2 coefficient can be written, up to the chosen transverse-basis orientation, as

$$\widehat{a}(k) = \frac{\gamma(k)}{|k|}\widehat{\omega}(k), \qquad |\gamma(k)| = 1, \tag{266}$$

where $\gamma(k)$ is the mode-dependent phase/sign factor generated by the basis convention. The four ablations isolate the two effects:

$$\boxed{F0:\ \widehat{\omega}, \qquad F1:\ \frac{\widehat{\omega}}{|k|}, \qquad F2:\ \gamma(k)\widehat{\omega}, \qquad F3:\ \frac{\gamma(k)}{|k|}\widehat{\omega} = \widehat{a}.} \tag{267}$$

Because $|\gamma(k)| = 1$,

$$\boxed{L_{F0} = L_{F2} = \sum_k |e_\omega(k)|^2, \qquad L_{F1} = L_{F3} = \sum_{k\neq 0} \frac{|e_\omega(k)|^2}{|k|^2} = L_u.} \tag{268}$$

Hence $1/|k|$ changes the spectral weighting, whereas the phase/sign factor changes the representation without changing its L^2 mode weights.

I Geometric Comparison of Three Evolution Representations

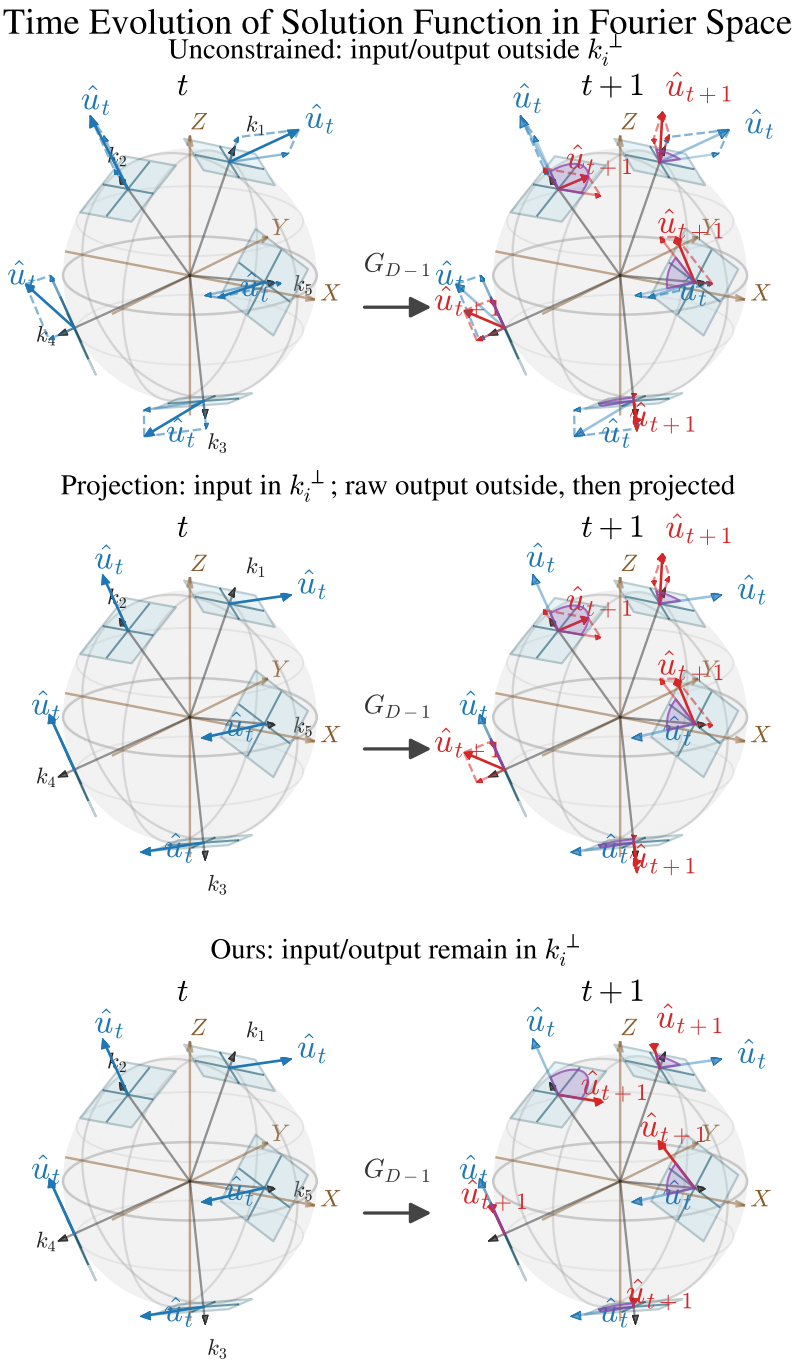


Figure 6: Geometric comparison of three temporal-evolution representations for each nonzero wavenumber k_i. **Top (unconstrained):** neither the input nor the output is required to be divergence-free or to lie in $k_i^\perp$, so the learned evolution uses ambient D-component modes. **Middle (Project-and-Generate):** the model uses an ambient D-component representation with no constraint on the longitudinal direction; its raw output is likewise unconstrained and is returned to $k_i^\perp$ only by a subsequent projection. **Bottom (ours):** the input and output are represented directly by the $D-1$ transverse coordinates in $k_i^\perp$. The neural operator never predicts the longitudinal component that projection would discard, and no projection is needed after prediction.

J REPRESENTATIVE PHYSICAL-SPACE RECONSTRUCTION GALLERY

Each figure compares the original field, its real-space $D-1$ latent channel(s), the reconstructed field, and pointwise error. The 3D figures additionally show xy, xz, and yz slices at 25%, 50%, and 75%.

J.1 WHY A REDUCED LATENT CAN EXTEND BEYOND A CLOSED PHYSICAL BOUNDARY

The white exterior of a closed-domain figure is not physical velocity data. To apply an FFT, the physical field is embedded in a periodic computational box $\widetilde{\Omega}$; write this array as $u_{\mathrm{emb}}(x,t) = \chi_\Omega(x)u(x,t)$, where χ_Ω is the physical-domain mask. Here "zero outside" is an embedding convention, not a claim that the physical velocity is defined outside the wall. The physical D-component input, reconstruction, and error are consequently masked outside Ω in the plots.

At each time t, the direct encoder on the nonzero modes is $a(\cdot,t) = \mathcal{F}_{D-1}^{-1}E^*\mathcal{F}_D u_{\mathrm{emb}}(\cdot,t)$. Let $B_{\ell j}(k) = [E(k)^*]_{\ell j} = E_{j\ell}(k)$ and define its continuous inverse-Fourier kernel by

$$K_{\ell j}(x) = \mathcal{F}_x^{-1}[B_{\ell j}(k)]\,(x), \qquad \mathcal{K}_{\ell j}(x,\tau) = K_{\ell j}(x)\,\delta(\tau). \tag{269}$$

Consequently the complete transform is a matrix-valued $D \to D-1$ convolution over space–time,

$$a_\ell(x,t) = \sum_{j=1}^{D}\int_{\mathbb{R}}\int_{\widetilde{\Omega}} \mathcal{K}_{\ell j}(x-y,t-s)\,u_{\mathrm{emb},j}(y,s)\,dy\,ds = \sum_{j=1}^{D}\int_{\widetilde{\Omega}} K_{\ell j}(x-y)\chi_\Omega(y)u_j(y,t)\,dy. \tag{270}$$

The $\delta(t-s)$ factor records that the fixed Fourier/Householder transform does not mix different times; its nonlocality is spatial. The zero Fourier mode remains the separately stored D-component mean described above. Thus, for $x \notin \Omega$, the interior values $y \in \Omega$ in Eq. (270) can yield $a_\ell(x,t) \neq 0$. These exterior values are valid reduced coordinates required by the exact transform; they are not a physical velocity outside the wall, and masking or deleting them would generally destroy exact decoding.

The degree of globality—the spatial extent and tail length of K—is controlled by the regularity of the Fourier multiplier $B(k)$. A smooth multiplier has rapidly decaying inverse-Fourier coefficients. Conversely, a sharper transition, larger discontinuity, or gauge seam in $E(k)$ produces more slowly decaying, broader spatial tails in K and therefore a more global convolution. The canonical Householder choice and the paired $k/-k$ convention can introduce such a directional seam (Appendix B); this explains why the latent exterior can be visibly global even when the physical field is restricted to a closed geometry. On a finite FFT box the kernel is periodized, so "local" means concentration and decay within that box, not compact support. This is distinct from the open-flow cards: there the virtual region is an intentional Fourier continuation of the physical inlet/outlet domain.

Why the boundary-exterior reduced latent is often more pronounced in three dimensions. For a two-dimensional physical field, the wavevector direction lies on S^1, whose tangent bundle admits a globally continuous unit tangent vector: for $n = (n_1, n_2)$ one may take $e(n) = (-n_2, n_1)^\top$. Thus the $2 \to 1$ transverse multiplier can be chosen globally smooth on the direction circle, and its inverse-Fourier kernel can be comparatively concentrated. In contrast, for a three-dimensional physical field the direction lies on S^2 and the $3 \to 2$ construction requires two independent tangent directions at every point. No single continuous global tangent frame exists on S^2 (the hairy-ball obstruction). Consequently, every one-chart $3 \to 2$ frame must have a directional discontinuity or singular chart transition. In the Householder construction this unavoidable defect appears near the reference-axis pole; changing the gauge can move the pole or seam, but cannot remove it.

This forced loss of angular regularity makes the three-dimensional multiplier $E(k)$ less smooth in Fourier space than a globally smooth two-dimensional choice. Its inverse-Fourier matrix kernel therefore has slower-decaying, more global spatial tails. Applied through Eq. (270), those longer tails carry interior field values farther into $x \notin \Omega$, so the $D-1$ latent outside a closed physical boundary is commonly much more visibly nonzero in three dimensions than in two. The exact magnitude still depends on the geometry and mask, the field spectrum near the seam, the FFT box, and the gauge. Moreover, the paired $k/-k$ convention used to make the latent real can add a coordinate seam even in two dimensions; the key distinction is that only the three-dimensional discontinuity is topologically unavoidable.

J.2 Open-boundary Fourier extension: accuracy and stored-coefficient trade-off

For a nonperiodic channel with inflow and outflow profiles that differ locally, a direct zero-padded FFT is not an admissible pure transverse representation: the boundary flux creates a longitudinal component. We instead construct a periodic, spectrally divergence-free extension U that restricts to the physical field, $RU \approx u$. The extension length is $g = \widetilde{L}/L$, and the axial Fourier density is $r = N_x^{\mathrm{Fourier}}/N_x$. At $g = 1.2$ and $r = 2$, the extension introduces only a 20% virtual interval while doubling the axial Fourier coefficient density. The resulting full codec payload stores exactly $D-1$ transverse coefficients at every nonzero mode (and the separately handled mean), yet reconstructs the physical field after decoding and cropping with relative error 8.45×10^{-15}.

Table 5 records the subsequent coefficient-sparsity experiment on this stable open-channel representation. The smallest Fourier coefficients were discarded by magnitude and the retained $D-1$ payload was decoded without an auxiliary longitudinal residual channel. The storage measure is the real-scalar-equivalent coefficient count divided by the original physical D-component array count; it does not charge index or quantization metadata. Thus the table directly separates the exact intrinsic component reduction from the practical cost of the extra Fourier modes required by an open-boundary extension.

Retained $D-1$ payload / original data	$\|RU_{\mathrm{decoded}} - u\|_2/\|u\|_2$	Interpretation
1.20 (full payload)	8.45×10^{-15}	machine-precision crop reconstruction
1.18	9.28×10^{-6}	below 10^{-5}, but exceeds the original storage
1.10	3.62×10^{-5}	10^{-5}–10^{-4} accuracy range
1.00	6.57×10^{-5}	10^{-5}–10^{-4} accuracy range at equal storage
0.90	9.46×10^{-5}	10^{-4} accuracy with 10% fewer stored values
0.50	3.01×10^{-4}	stronger compression with reduced accuracy

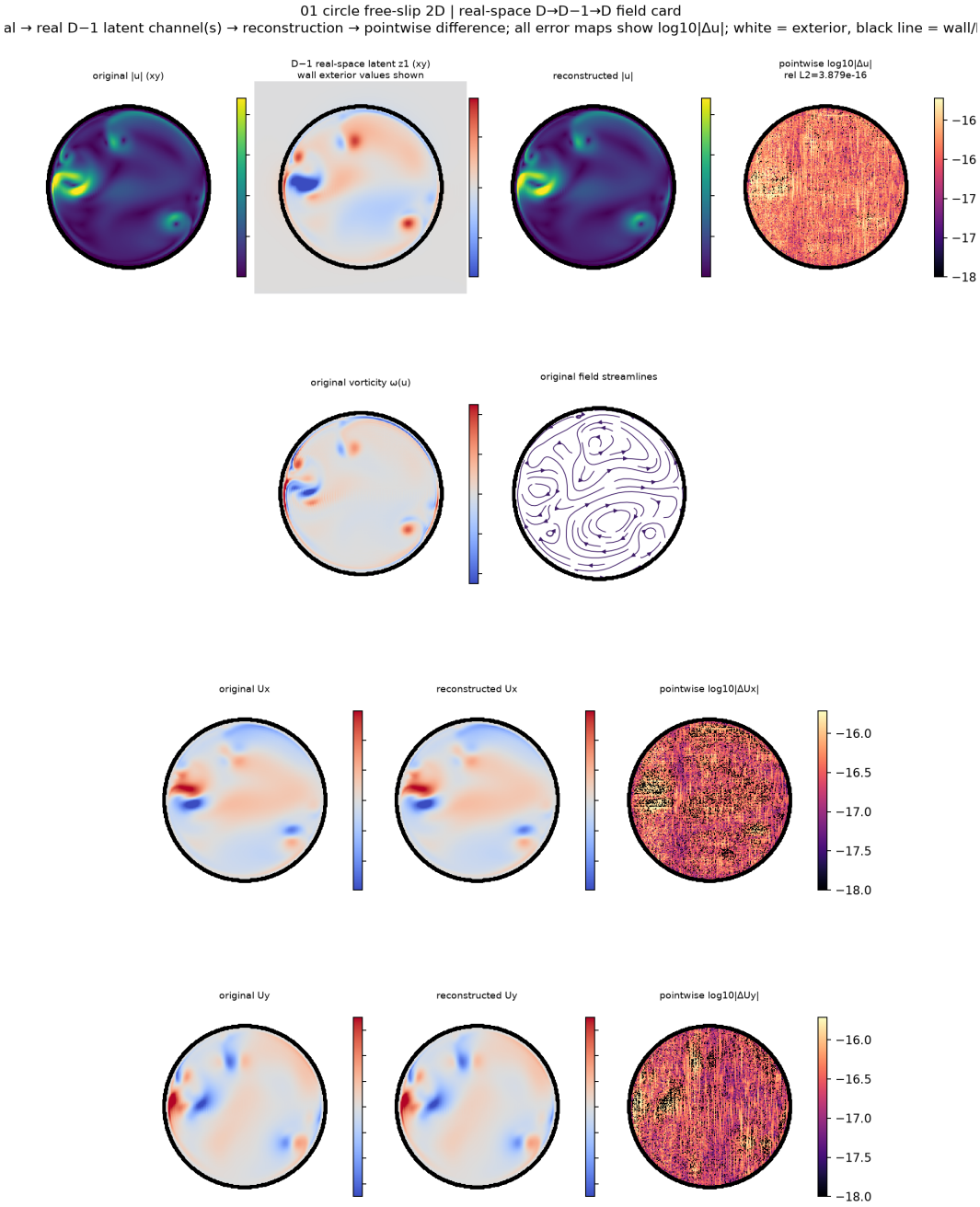
Table 5: Open-channel Fourier-extension storage–accuracy frontier. The extension uses $g = 1.2$ and axial density $r = 2$ before coefficient pruning. Every nonzero retained mode uses only the intrinsic $D-1$ transverse coordinates; no longitudinal residual is stored.

These results establish two positive facts. First, different inflow and outflow profiles do not prevent high-accuracy $D \to D-1 \to D$ reconstruction: after sufficient Fourier extension and spectral oversampling, the strict periodic divergence-free extension is recovered to machine precision. Second, the same representation can be truncated below the original total data count while retaining a relative physical-domain error of order 10^{-4} (and the 10^{-5}–10^{-4} range at equal storage). For this single-field, magnitude-pruning experiment, a strict error below 10^{-5} required 1.18 times the original storage. Hence per-mode $D \to D-1$ reduction is exact, while a guarantee of both sub-10^{-5} error and lower total storage would require additional data-dependent compression structure beyond independent Fourier-coefficient pruning.

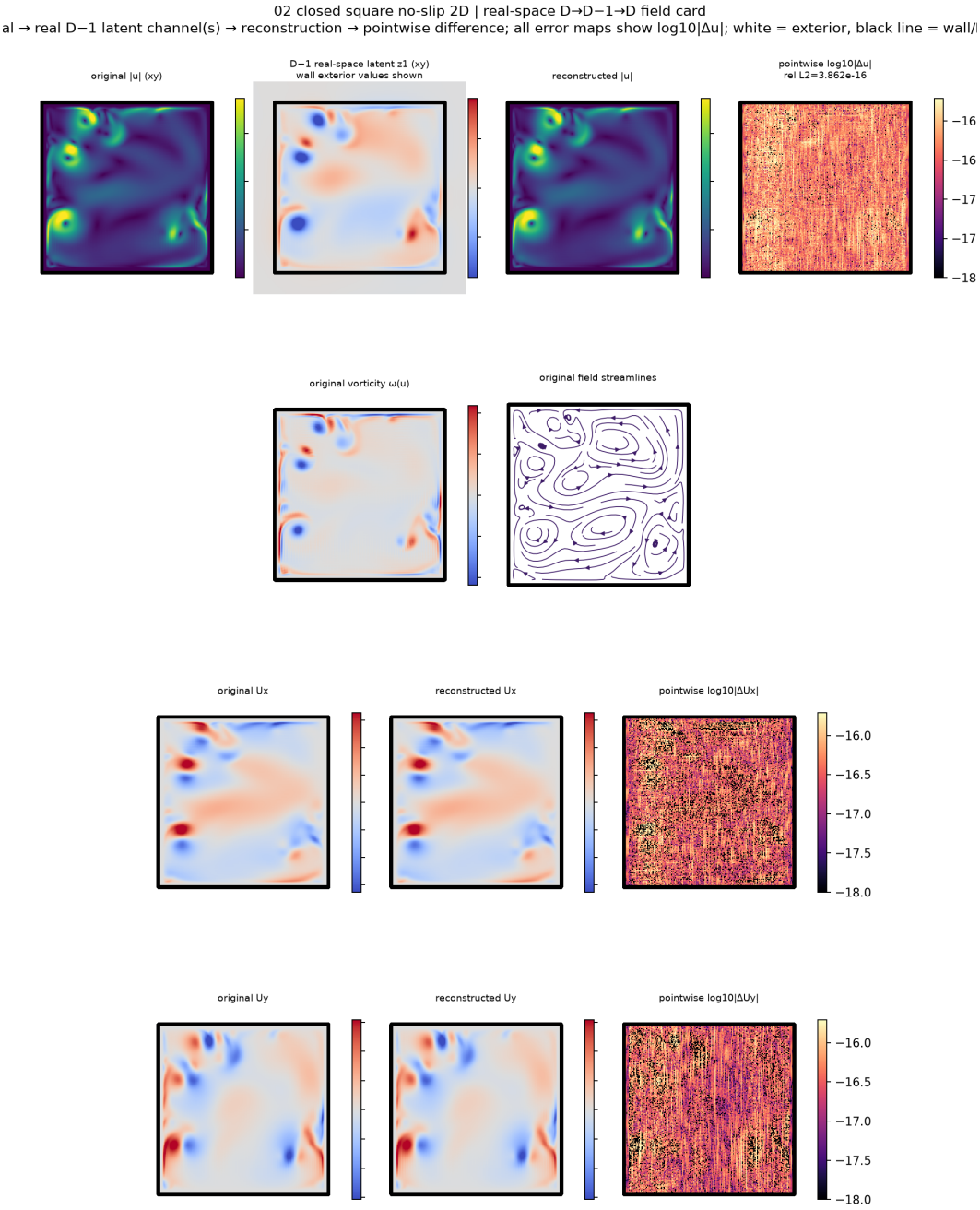
Why the early open-boundary reconstructions were inaccurate. The initial low-density Fourier extensions produced percent-level physical-domain discrepancies (often around 1%–2%) and, in the poorly resolved virtual region, could generate very large amplitudes. This was not a failure of the $D \to D-1$ codec: encoding and decoding an already admissible periodic solenoidal extension is lossless to roundoff. The issue was that a coarse Fourier representation could not simultaneously resolve the continuation and enforce the required constraints robustly. Increasing the axial Fourier-coefficient density supplied the missing degrees of freedom; with the exact solenoidal constraint imposed on this refined representation, the virtual-region amplitude remained controlled and the physical crop converged to machine precision. Thus Fourier refinement is a numerical-stability requirement of the extension construction, not an extra learned channel or a change to the intrinsic $D-1$ representation.

Spectral versus finite-difference divergence. For discretized data, "divergence-free" is defined relative to the differentiation operator used to test it. A field can satisfy $D_p^{\mathrm{FD}} \cdot u \approx 0$ for a finite-difference stencil of order p while failing the FFT condition $\mathcal{F}^{-1}[ik \cdot \widehat{u}] = 0$ required by the trans-

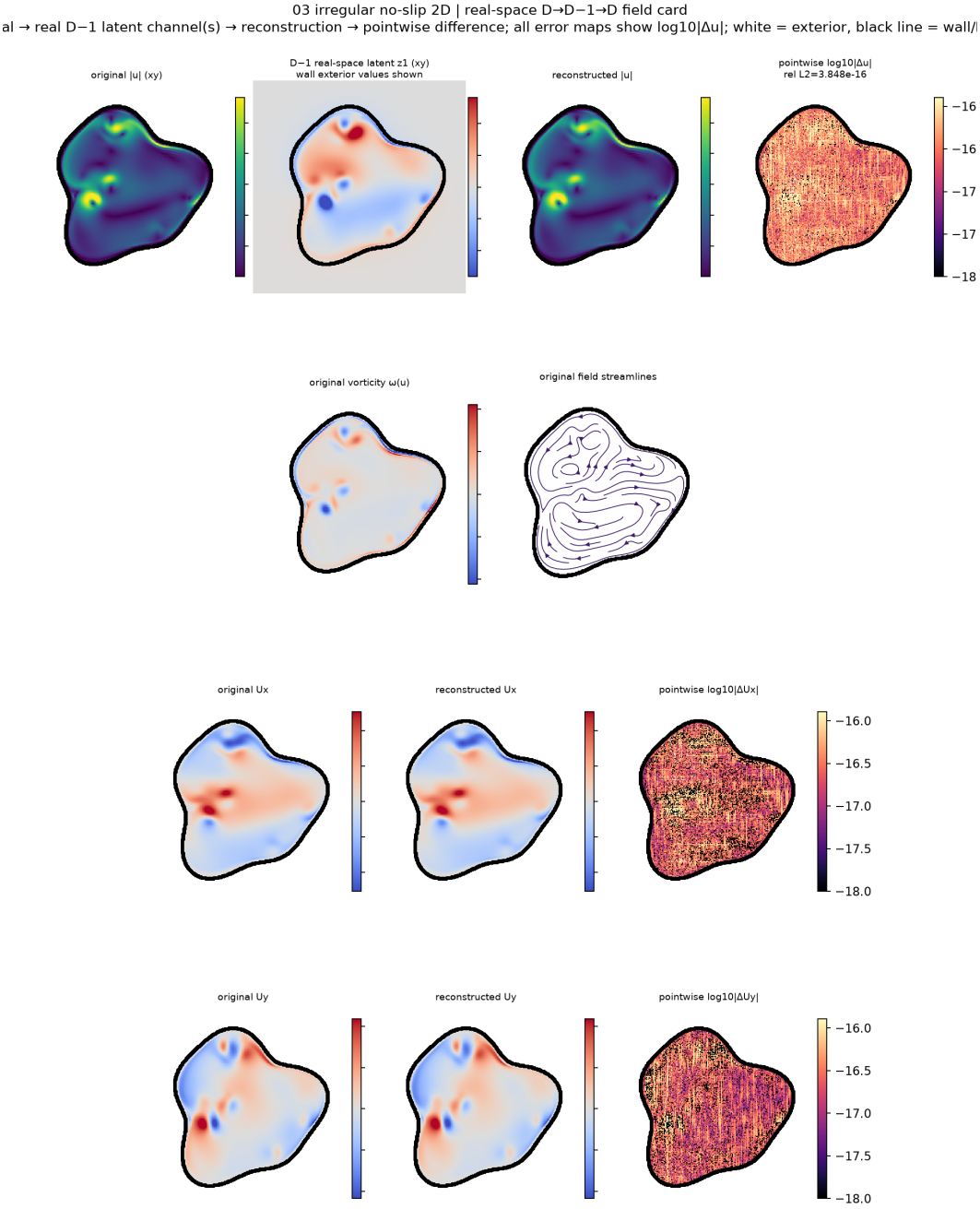
verse Fourier codec, because the finite-difference multiplier $\widetilde{k}_p(kh)$ is generally not the spectral multiplier k. Low-order stencils can therefore report a small finite-difference divergence while leaving appreciable longitudinal Fourier content. In the derivative-order diagnostic, raising the centered stencil order through 8, 10, and 12 made the finite-difference diagnostic progressively closer to the spectral result on resolved interior modes. It remains an approximation at any finite order, however; strict spectral divergence-freeness is established only by the Fourier-space condition itself (or by a representation built for the solver's own discrete symbol).



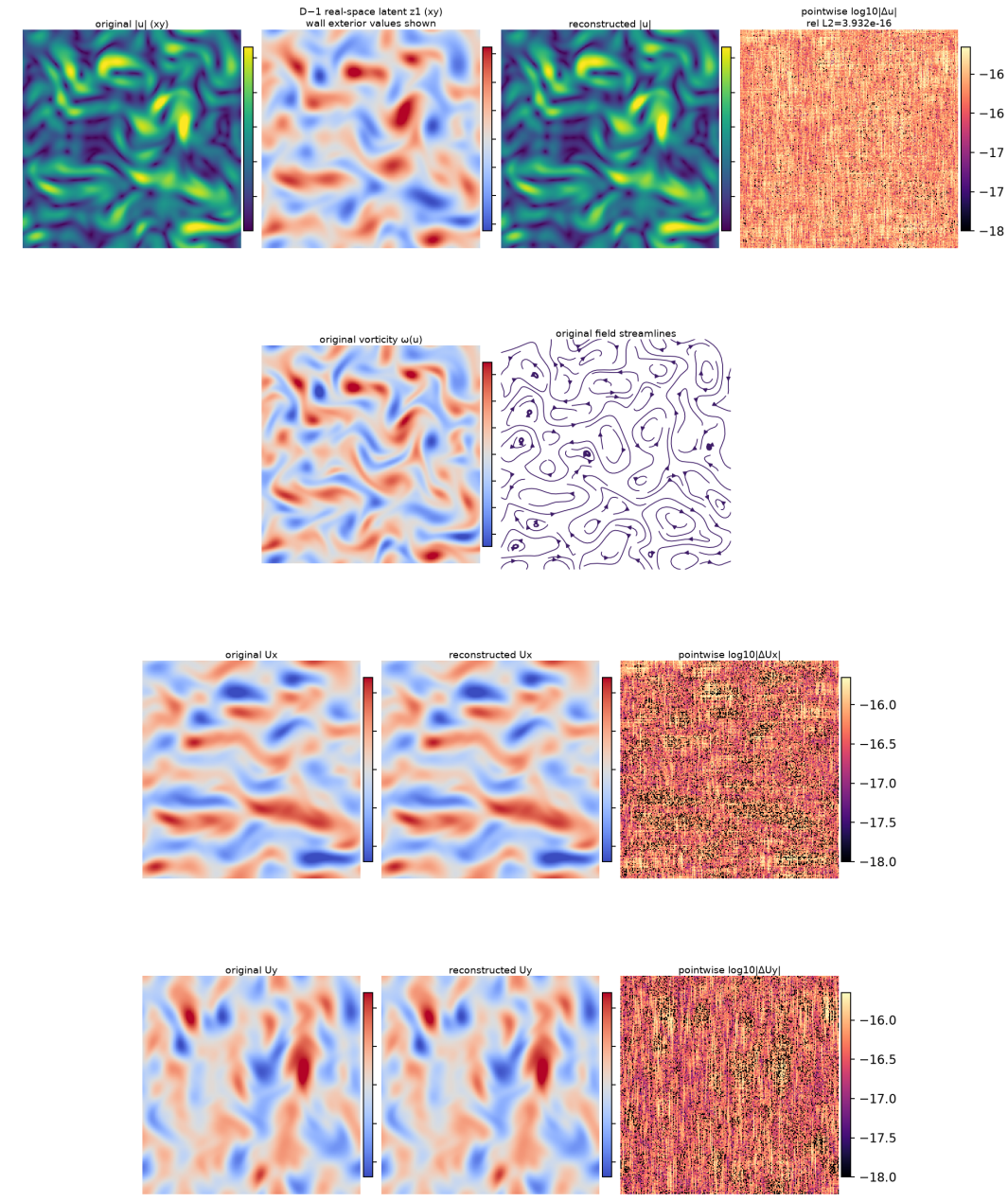
2D circular free-slip cavity. This is a two-dimensional disk. Its circumference is a closed free-slip and impermeable boundary ($u \cdot n = 0$), so it has neither periodic identification nor inlet/outlet Fourier extension. Outside the black physical boundary, the D-component physical field and displayed reconstruction are deliberately masked; the $D - 1$ latent is deliberately left visible because the Fourier/Householder encoder is the nonlocal spatial convolution in Appendix J.1.



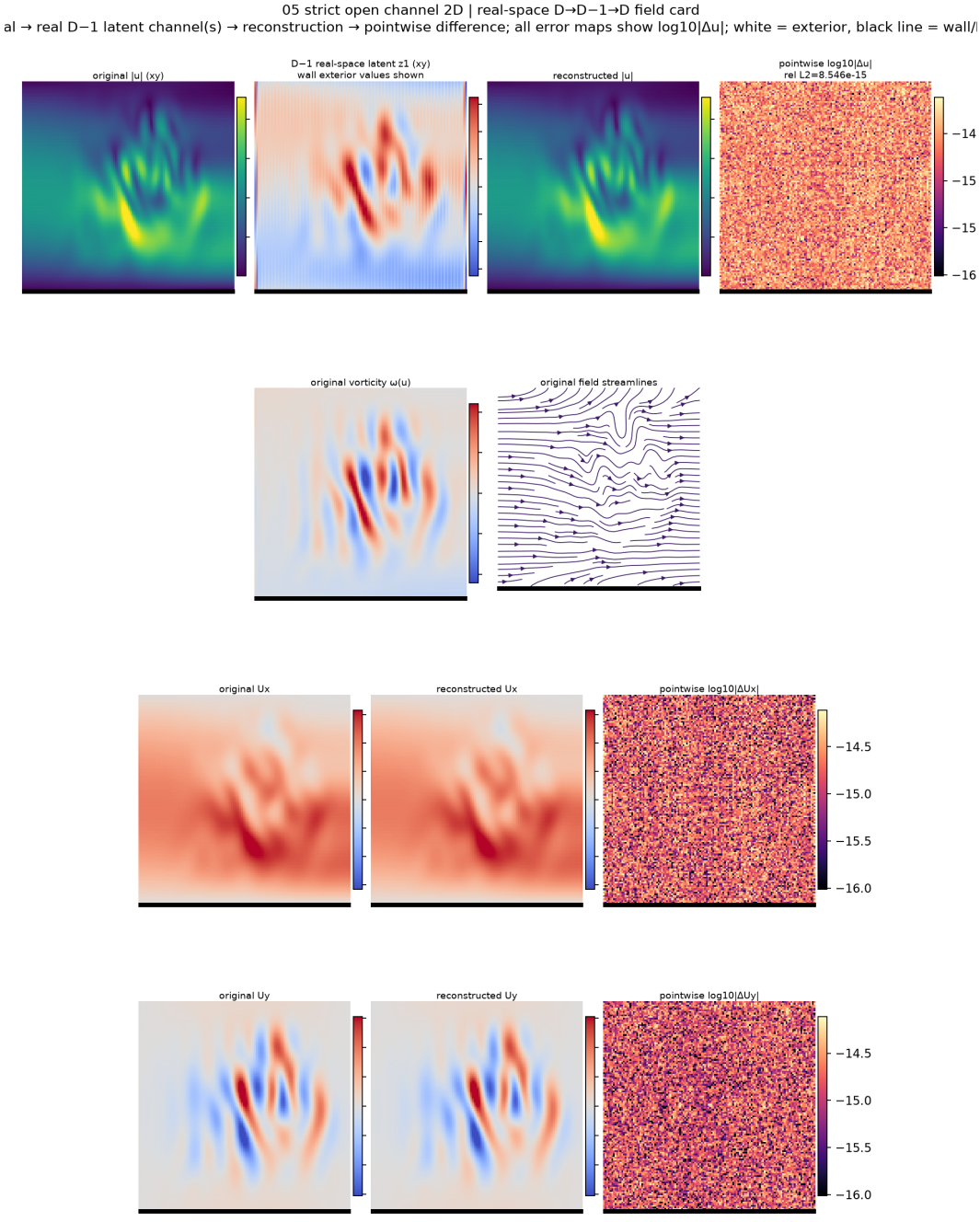
2D closed square with no-slip walls. This is a two-dimensional square cavity. All four sides are closed, impermeable no-slip walls; it is not periodic and has no inflow or outflow. Outside the black physical boundary, the D-component physical field and displayed reconstruction are deliberately masked; the $D-1$ latent is deliberately left visible because the Fourier/Householder encoder is the nonlocal spatial convolution in Appendix J.1.



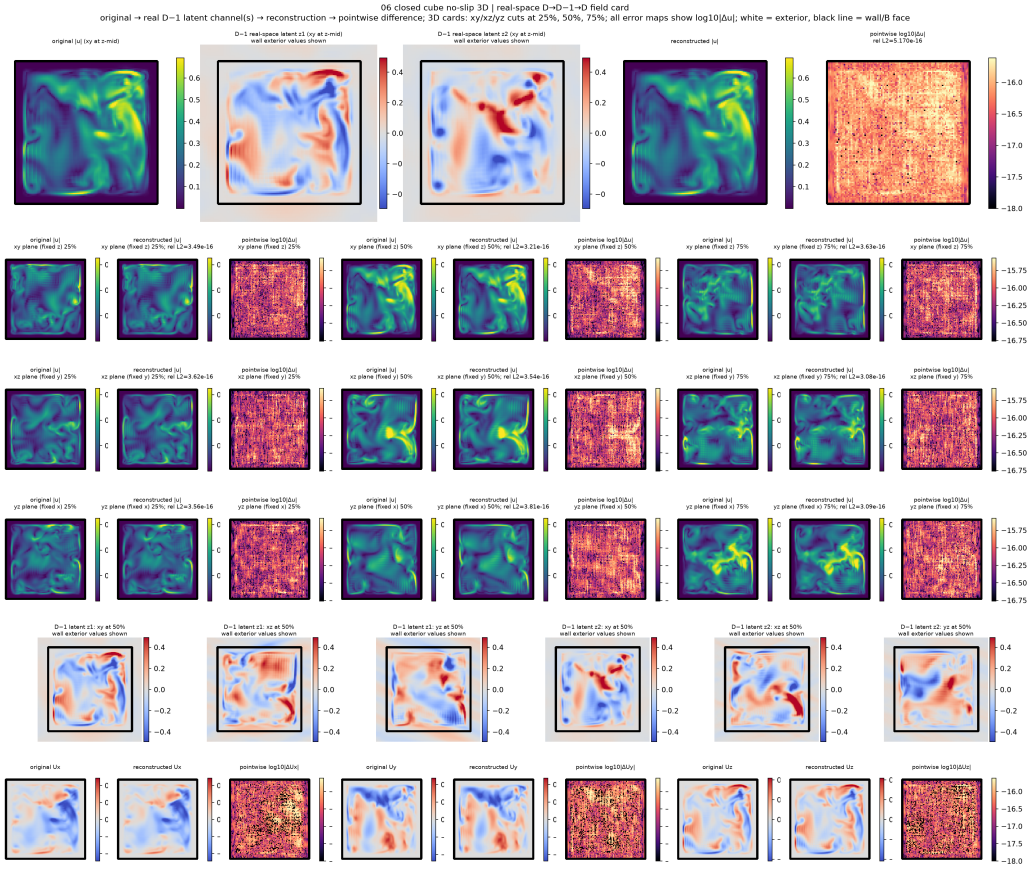
2D irregular closed no-slip domain. This two-dimensional simply connected irregular domain is bounded by a solid no-slip wall. The exterior is masked white, and there is no periodic matching or open-boundary extension. Outside the black physical boundary, the D-component physical field and displayed reconstruction are deliberately masked; the $D-1$ latent is deliberately left visible because the Fourier/Householder encoder is the nonlocal spatial convolution in Appendix J.1.



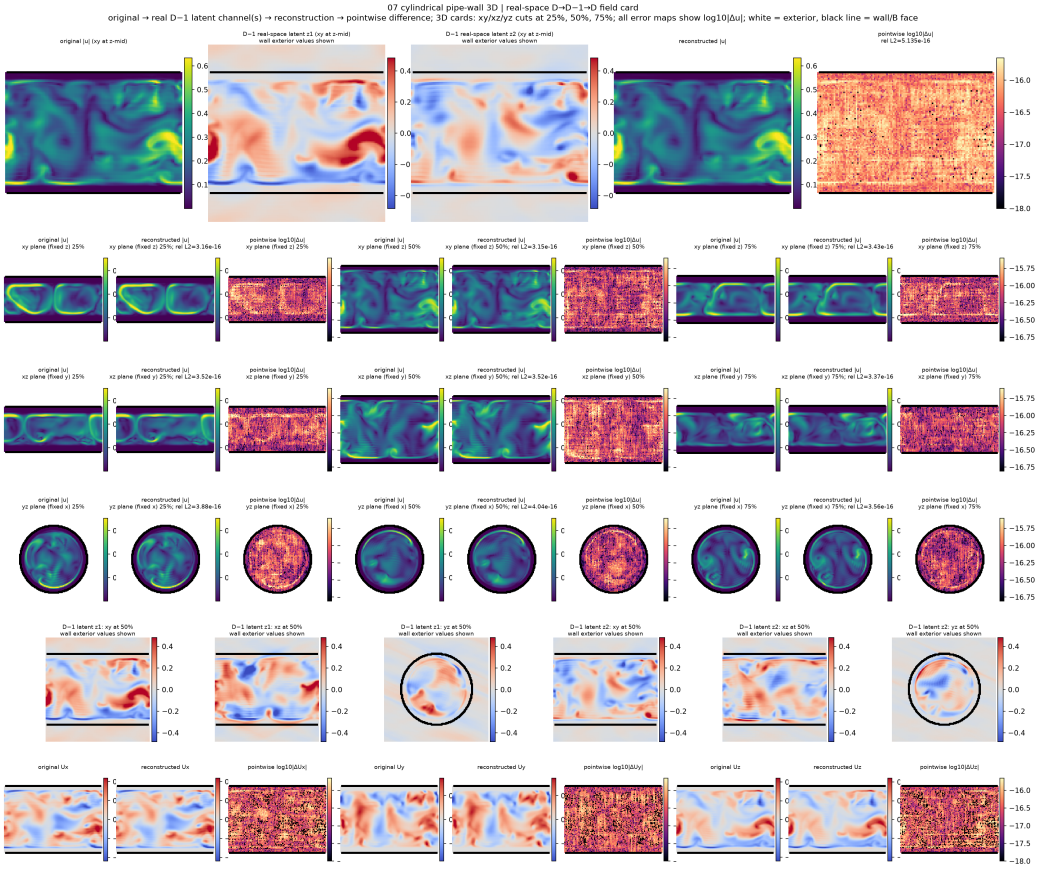
2D periodic square turbulence. This is a two-dimensional square with opposite sides identified, i.e., a periodic torus. It has no physical wall or inlet/outlet boundary, so no Fourier continuation is required.



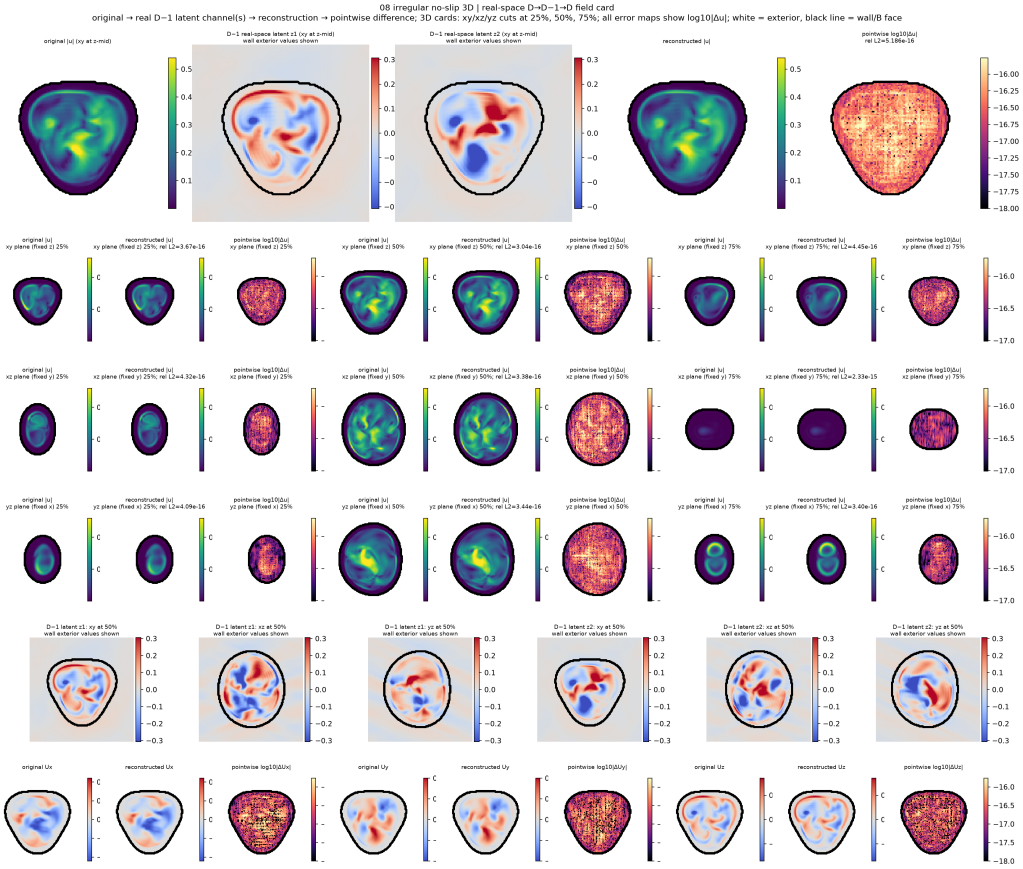
2D strict open channel. This is a two-dimensional channel with nonperiodic inlet/outlet ends. Its Fourier representation therefore uses an extended periodic box rather than imposing direct periodicity on the physical open channel. The virtual portion is an intentional Fourier continuation, not an exterior latent tail of a closed-wall field.



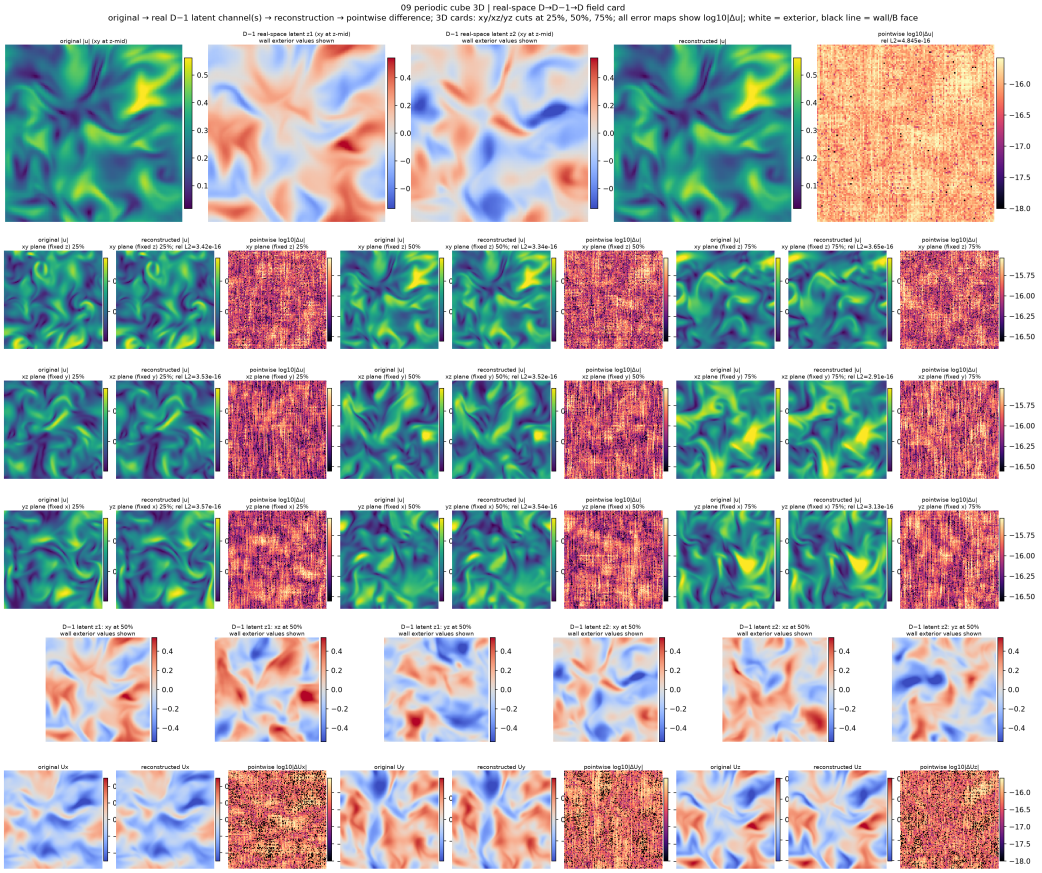
3D closed cube with no-slip walls. This is a three-dimensional cubical cavity. Its faces are closed no-slip and impermeable walls, and the xy, xz, and yz slices expose the enclosed geometry. Outside the black physical boundary, the D-component physical field and displayed reconstruction are deliberately masked; the $D-1$ latent is deliberately left visible because the Fourier/Householder encoder is the nonlocal spatial convolution in Appendix J.1.



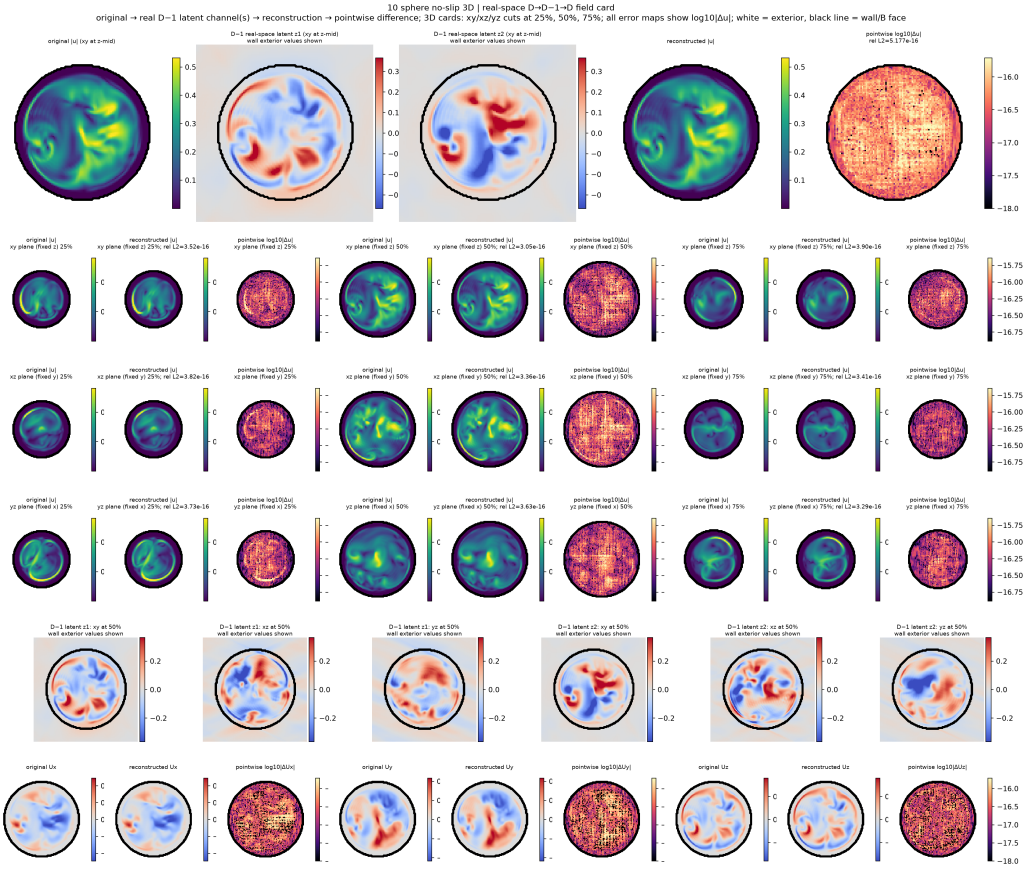
3D cylindrical pipe with side wall. This is a three-dimensional circular pipe. The axial end faces are inlet/outlet openings while the cylindrical side wall is solid; a nonperiodic axial treatment requires Fourier extension. Outside the solid side wall, the D-component physical field and displayed reconstruction are deliberately masked, while the $D-1$ latent is left visible because the Fourier/Householder encoder is the nonlocal spatial convolution in Appendix J.1.



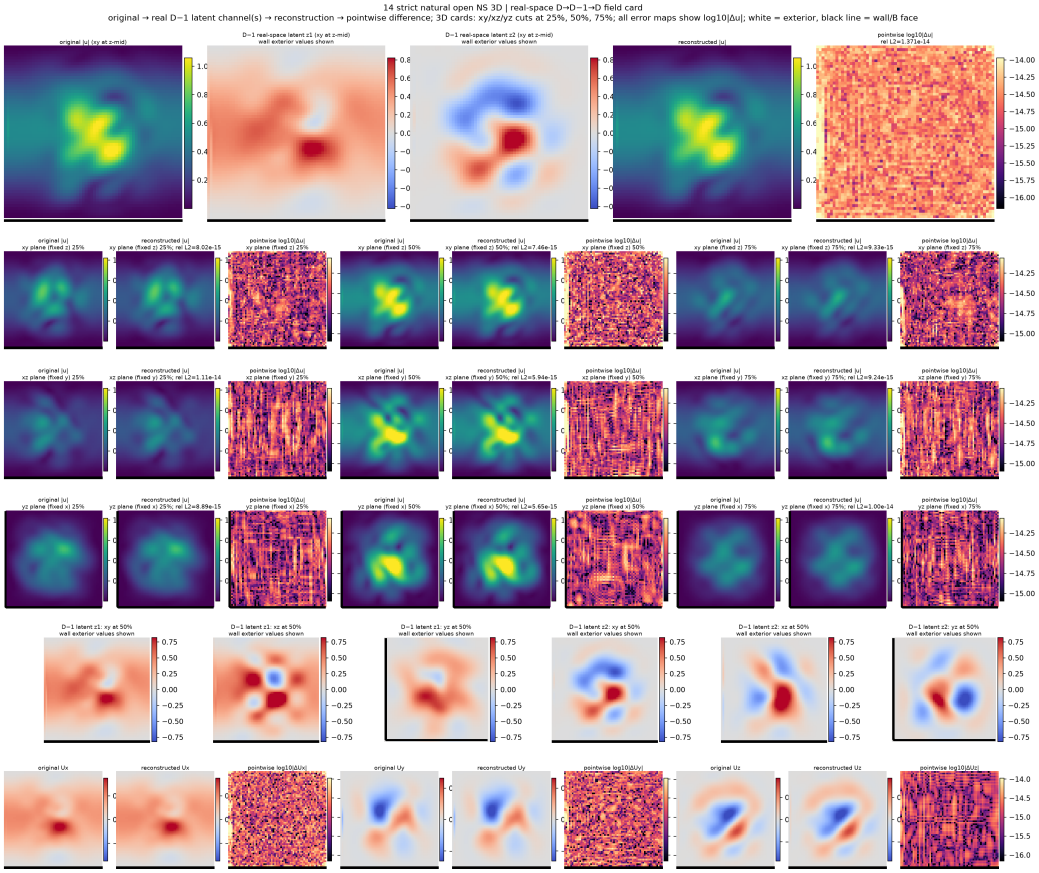
3D irregular closed no-slip domain. This is a three-dimensional irregular solid-bounded volume. The black contours mark the closed, impermeable no-slip boundary, while white regions lie outside the physical domain. Outside the black physical boundary, the D-component physical field and displayed reconstruction are deliberately masked; the $D-1$ latent is deliberately left visible because the Fourier/Householder encoder is the nonlocal spatial convolution in Appendix J.1.



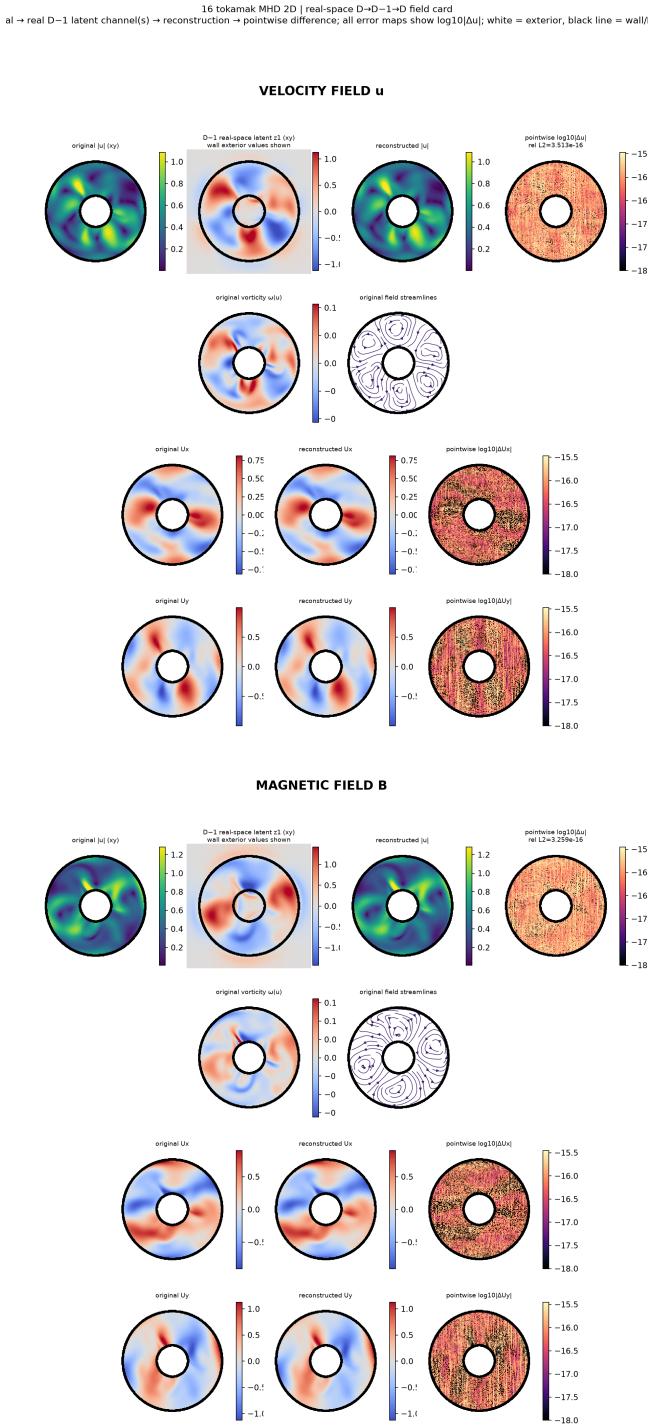
3D periodic cube. This is a three-dimensional periodic box with opposite faces identified. It has no physical wall or inlet/outlet boundary and is represented directly in Fourier space.



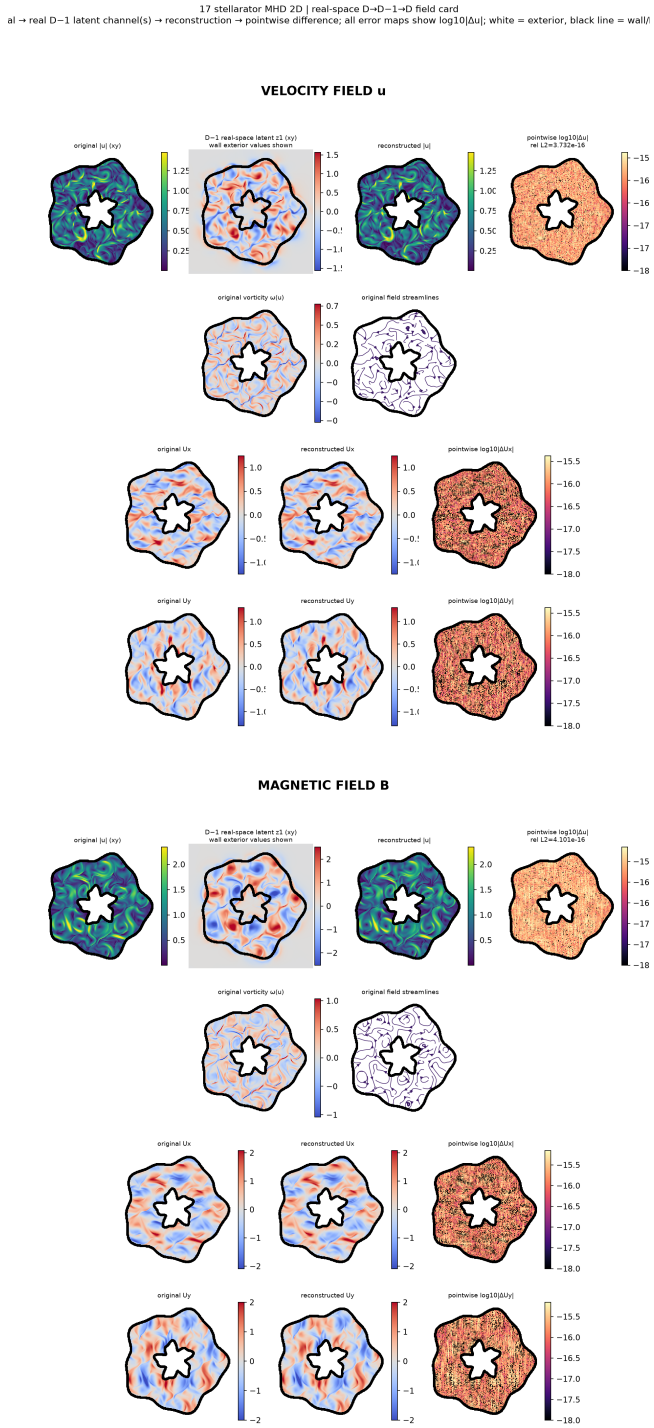
3D spherical no-slip cavity. This is a three-dimensional spherical enclosed domain. Its spherical surface is a closed no-slip, impermeable wall rather than a periodic or open boundary. Outside the black physical boundary, the D-component physical field and displayed reconstruction are deliberately masked; the $D-1$ latent is deliberately left visible because the Fourier/Householder encoder is the nonlocal spatial convolution in Appendix J.1.



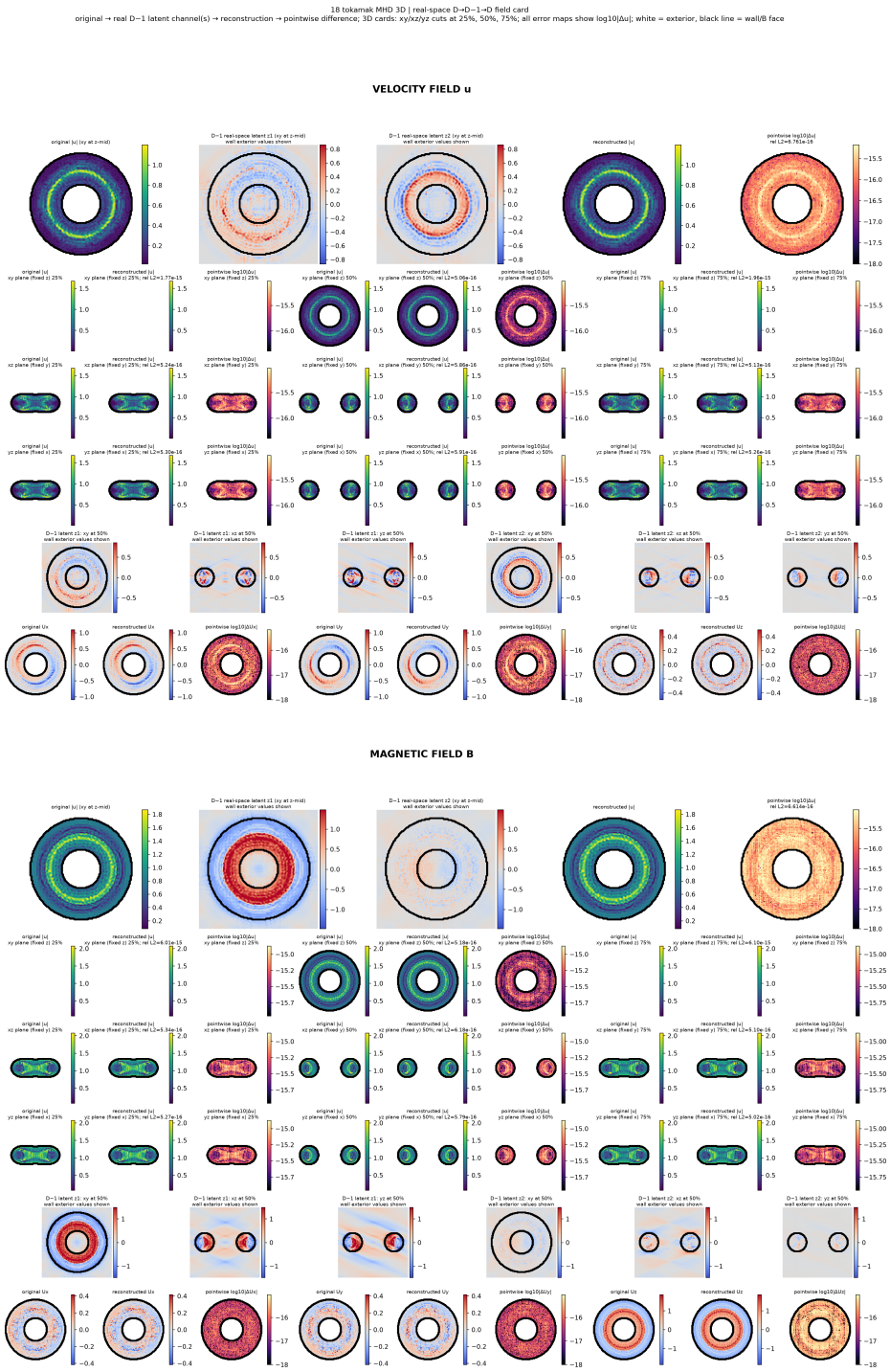
3D natural-open Navier–Stokes field. This is a three-dimensional nonperiodic open-flow state. The open axial direction is encoded through Fourier extension to an enlarged periodic box rather than by treating the physical domain as periodic. Its virtual box region is part of that intentional continuation, not a closed-boundary exterior latent.



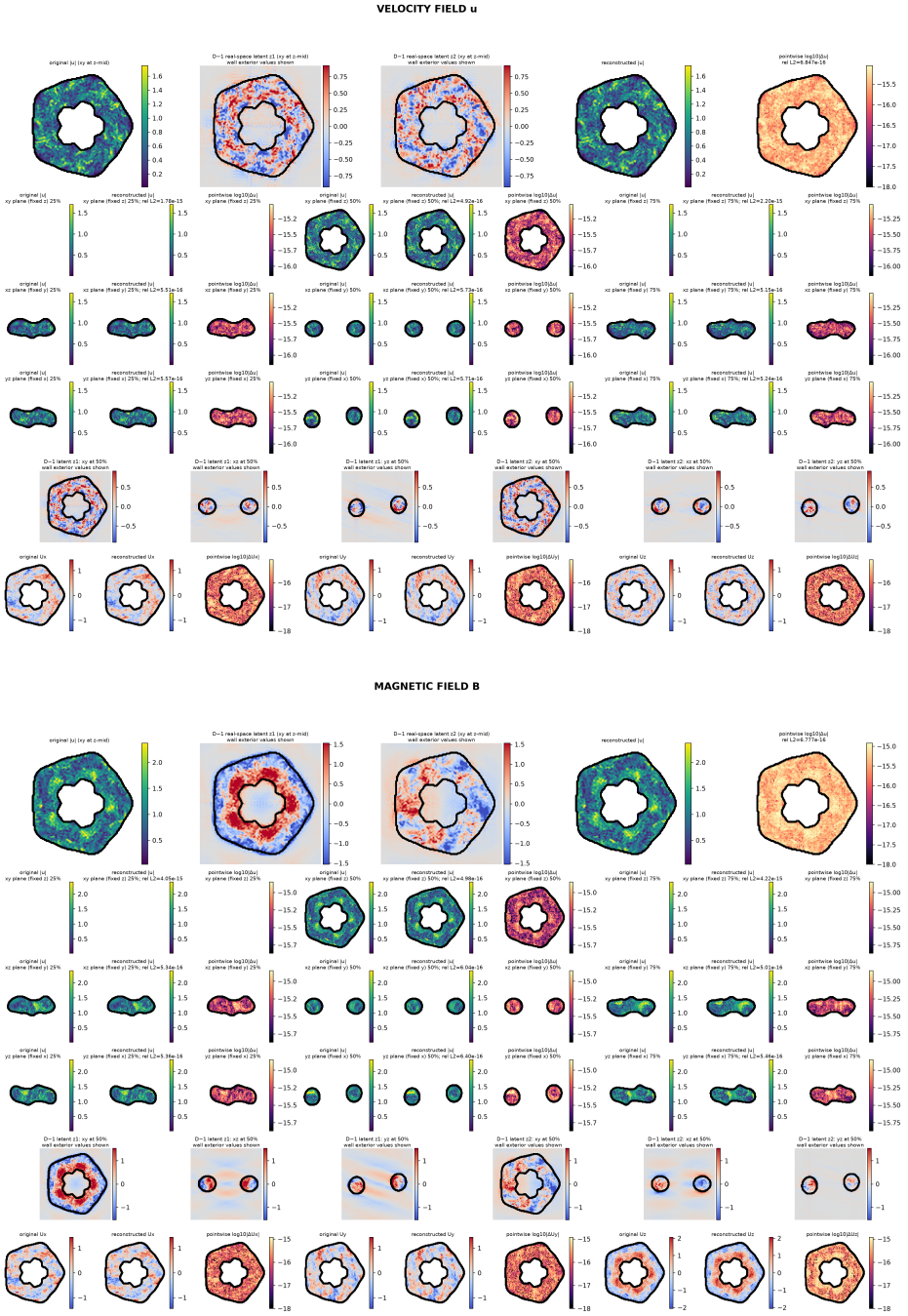
2D tokamak MHD cross-section. This figure contains both velocity u and magnetic field B on a bounded annular tokamak cross-section. The black curves mark the closed physical boundary in the displayed plane; it is not an inlet/outlet extension setting. Outside the black physical boundary, the D-component physical field and displayed reconstruction are deliberately masked; the $D-1$ latent is deliberately left visible because the Fourier/Householder encoder is the nonlocal spatial convolution in Appendix J.1.



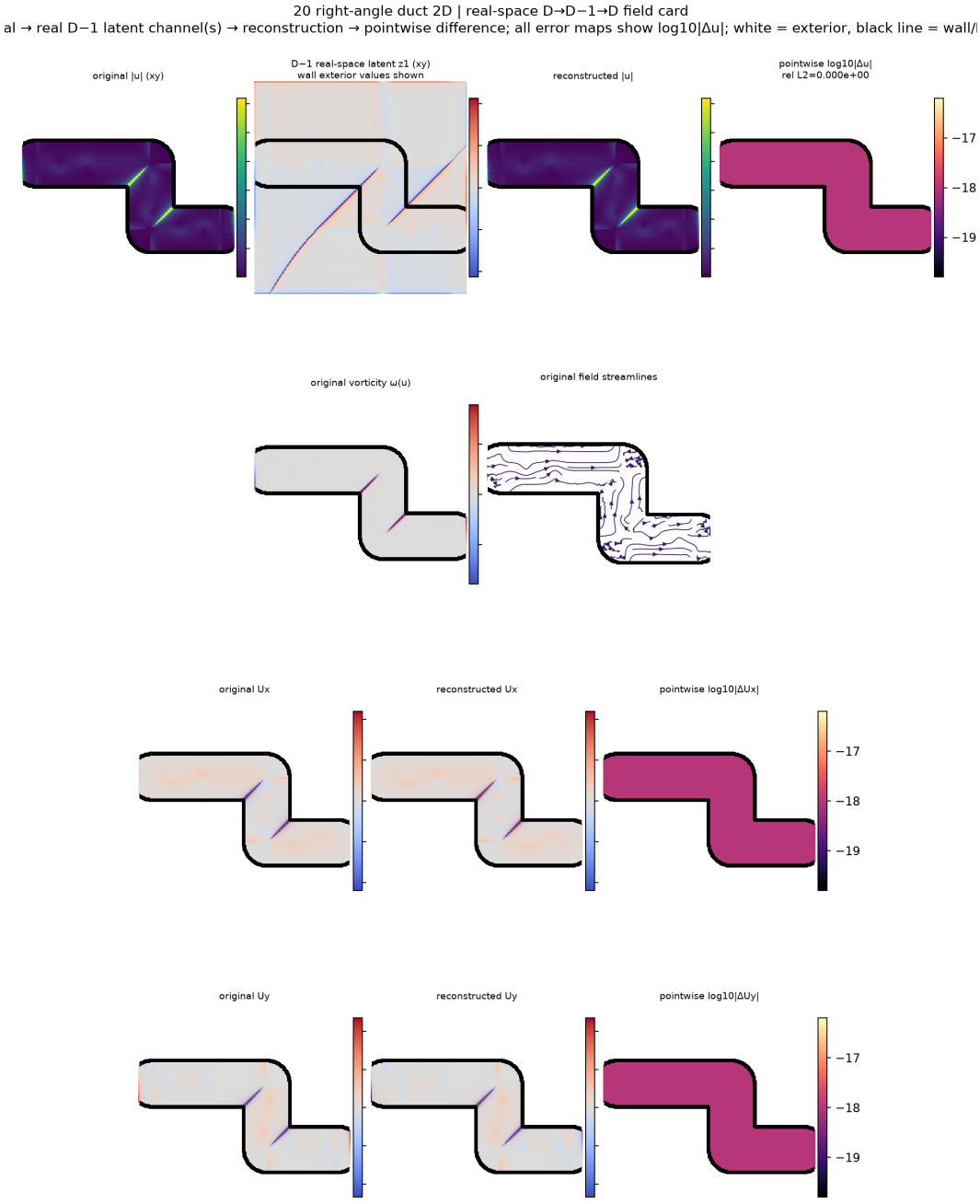
2D stellarator MHD cross-section. This figure contains both u and B on an irregular annular stellarator cross-section. The physical cross-section is wall bounded and closed, not a directly periodic or open-channel geometry. Outside the black physical boundary, the D-component physical field and displayed reconstruction are deliberately masked; the $D-1$ latent is deliberately left visible because the Fourier/Householder encoder is the nonlocal spatial convolution in Appendix J.1.



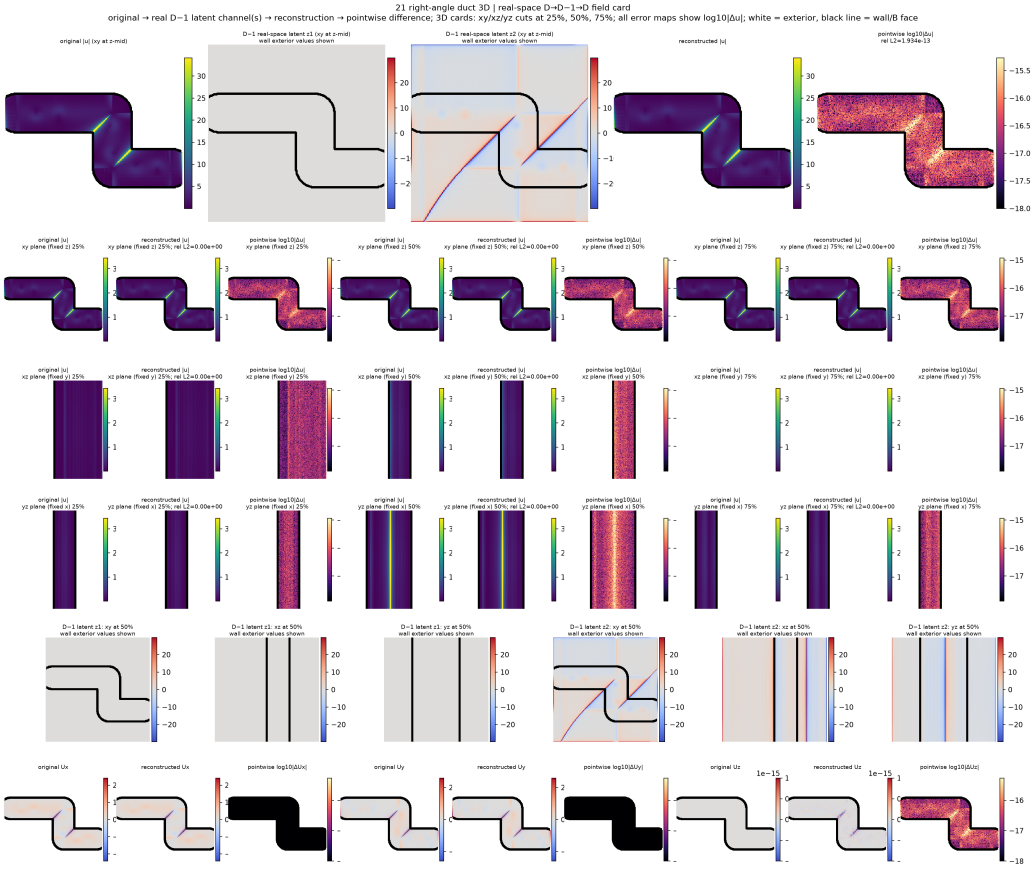
3D tokamak MHD configuration. This is a three-dimensional toroidal MHD configuration with velocity and magnetic fields. The toroidal coordinate is periodic, whereas the cross-sectional plasma/wall boundary is closed; no inlet/outlet extension is used. Outside the black physical boundary, the D-component physical field and displayed reconstruction are deliberately masked; the $D-1$ latent is deliberately left visible because the Fourier/Householder encoder is the nonlocal spatial convolution in Appendix J.1.



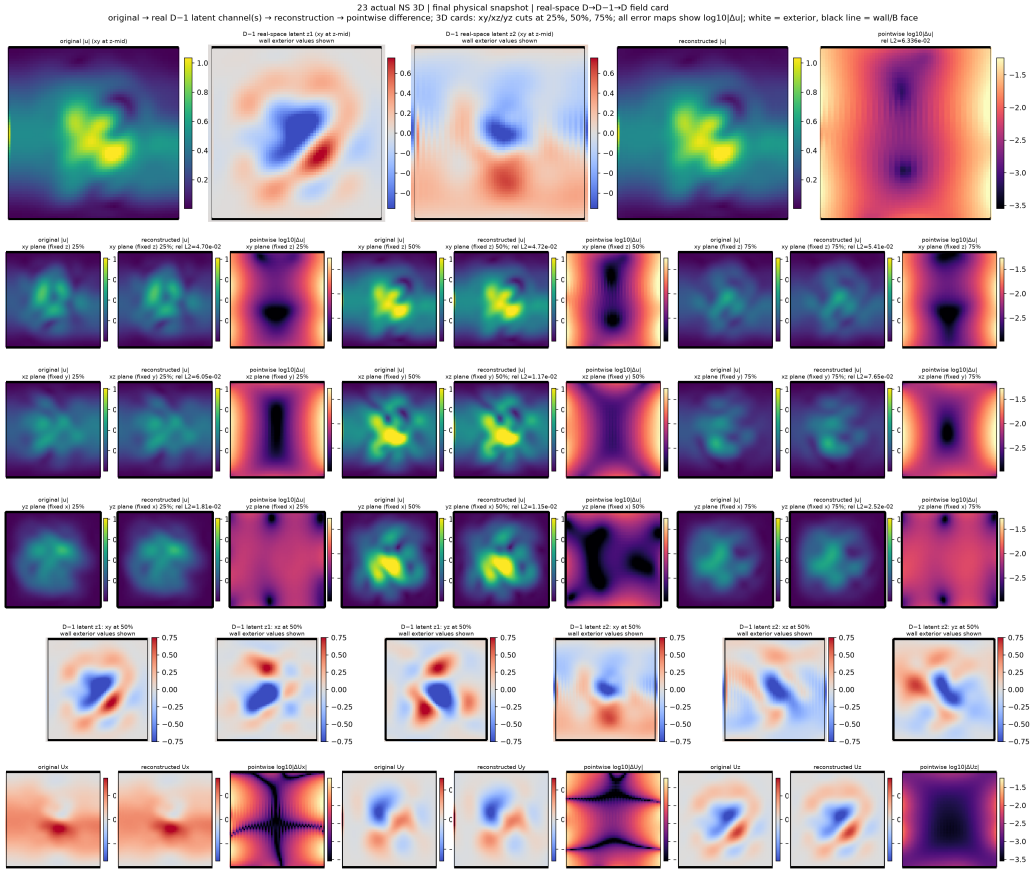
3D stellarator MHD configuration. This is a three-dimensional twisted toroidal stellarator configuration with both u and B. The toroidal direction is periodic while the irregular cross-sectional boundary is closed, not an open inflow/outflow boundary. Outside the black physical boundary, the D-component physical field and displayed reconstruction are deliberately masked; the $D-1$ latent is deliberately left visible because the Fourier/Householder encoder is the nonlocal spatial convolution in Appendix J.1.



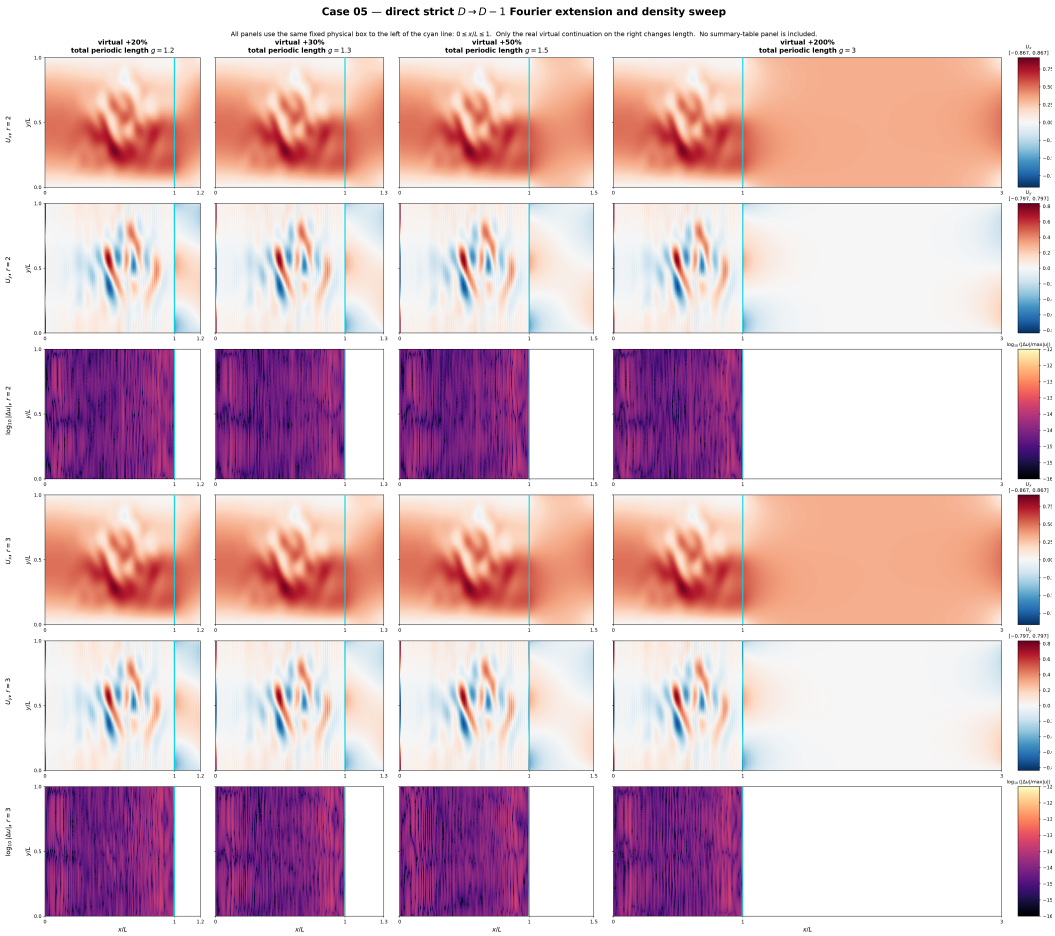
2D right-angle duct. This is a two-dimensional right-angle duct enclosed by the solid wall shown in black. It is a wall-bounded, nonperiodic geometry rather than a periodic square or an open-extension benchmark. Outside the black physical boundary, the D-component physical field and displayed reconstruction are deliberately masked; the $D-1$ latent is deliberately left visible because the Fourier/Householder encoder is the nonlocal spatial convolution in Appendix J.1.



3D right-angle duct. This is the three-dimensional counterpart of the right-angle duct. Its wall-bounded geometry is identified by the black contours across all slice orientations, with no periodic identification imposed. Outside the black physical boundary, the D-component physical field and displayed reconstruction are deliberately masked; the $D-1$ latent is deliberately left visible because the Fourier/Householder encoder is the nonlocal spatial convolution in Appendix J.1.



3D Navier–Stokes final snapshot. This is an actual three-dimensional Navier–Stokes physical snapshot.



Fourier-extension length and density sweep. This composite uses the 2D open channel and compares 20%, 30%, 50%, and 200% virtual extensions ($g = 1.2, 1.3, 1.5, 3.0$) at axial Fourier densities $r = 2$ and $r = 3$. The cyan divider separates the physical region from the intentional Fourier-extended virtual region.

K REDUNDANCY, OPTIMIZATION, AND IDENTIFIABILITY DIAGNOSTICS

This section records representation-learning diagnostics separately from the evolution experiments. The aim is to distinguish two facts that are often conflated: a redundant parameterization can contain directions that do not change the physical divergence-free field, and a particular optimization run can nevertheless be easier or harder for reasons that include architecture, conditioning, and optimization dynamics. The results below document both the observed optimization behavior and the exact null-space mechanism.

K.1 THREE-DIMENSIONAL BALL EVOLVING IN TIME: A $D = 4$ SPACE–TIME FIELD

The first study treats a three-dimensional ball over time as a four-coordinate space–time field (x, y, z, t) and learns its conserved current. Five parameterizations with approximately 26k trainable parameters were trained for 600 seconds each. Every parameterization used the same fixed stochastic collocation stream; only the ten random neural-network initializations were changed. The reported loss is the best fixed-validation loss attained by a run, while the final two columns make the optimization speed explicit.

Representation	rank / nullity	best validation loss	optimizer steps	loss decades / s
Minimal Fourier $a \in \mathbb{R}^3$ (ours)	$3/0$	$2.734\times10^{-2} \pm 2.619\times10^{-3}$	$16{,}106 \pm 80$	5.286×10^{-3}
Direct complex Fourier coefficients	$3/0$	$7.735 \pm 1.404\times 10^{-2}$	$16{,}205 \pm 71$	1.199×10^{-3}
Quaternion three-potential	$3/0$	$1.373\times10^{-3} \pm 1.977\times10^{-4}$	$28{,}306 \pm 909$	7.445×10^{-3}
NCL vector auxiliary field $B \in \mathbb{R}^4$	$3/1$	$2.516\times10^{-2} \pm 3.241\times10^{-3}$	$15{,}034 \pm 54$	5.400×10^{-3}
NCL antisymmetric potential $A \in \mathbb{R}^6$	$3/3$	$1.039\times10^{-3} \pm 6.298\times10^{-5}$	$28{,}351 \pm 233$	7.661×10^{-3}

Table 6: Ten-initialization representation-learning record for the three-dimensional ball over time. Mean $\pm$ standard deviation is reported for loss and optimizer steps. The rank and nullity are per nonzero $D = 4$ Fourier mode.

In this capacity-matched, fixed-budget record, the antisymmetric NCL potential reached the lowest validation loss and the fastest loss decrease among the listed alternatives. The paired runs also gave lower final validation loss for the NCL potential in all ten comparisons and a broader, flatter measured low-loss neighborhood. The following separate controlled sweep is the missing experiment that deliberately increases the redundancy of the potential representation.

K.2 CONTROLLED SWEEP WITH DELIBERATELY OVERCOMPLETE POTENTIAL COORDINATES

Here every model solves the same $D = 4$ Ball evolution problem for exactly $10{,}000$ updates, with the same sampler, optimizer, hidden architecture, initialization seed, and stochastic collocation stream. Only the learned potential coordinates and the fixed map from those coordinates to the physical current are changed. Thus the final validation loss after the common budget is the optimization-progress comparison, while the wall-clock column reports its computational cost. The direct antisymmetric baseline predicts the six independent entries of A. In the dense rows, an unconstrained vector $z \in \mathbb{R}^m$ is first mapped by a fixed dense co-isometry P to those six entries, $A = \mathrm{skew}(Pz)$. In the multi-potential rows, several independently predicted antisymmetric potentials are summed. In the outer-product rows, no network directly predicts an antisymmetric matrix: each branch predicts unrestricted vectors $p_r, q_r \in \mathbb{R}^4$ and the decoder constructs

$$A = \frac{1}{\sqrt{r/2}} \sum_{s=1}^{r} \left(p_s q_s^{\mathsf{T}} - q_s p_s^{\mathsf{T}}\right).$$

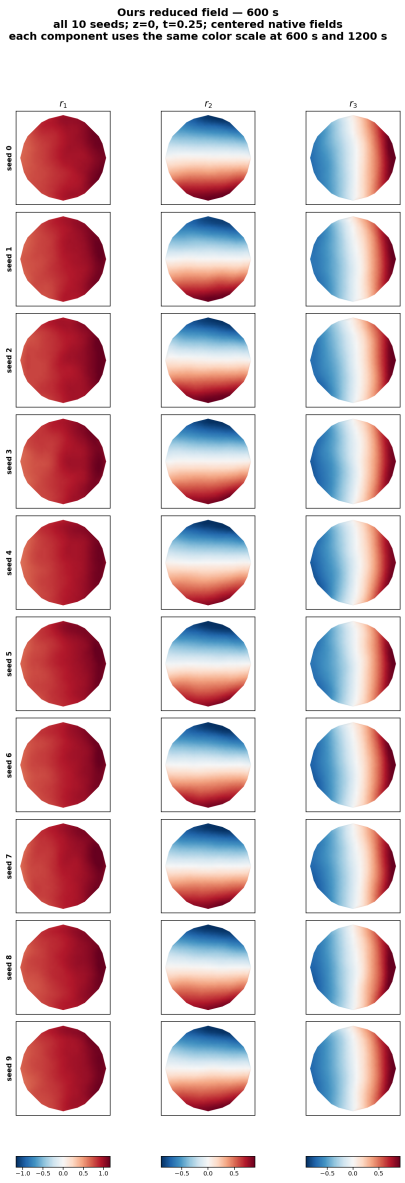
This is the unrestricted vector–outer-product construction in the original experiment record.

Potential coordinates	output coordinates	nullity	validation loss at 10k ↓	training s
Quaternion three-potential construction	3	0	0.1164051	111.21
Vector auxiliary field B	4	1	0.1784628	245.97
Direct antisymmetric potential A	6	3	0.1038496	117.72
Canonical skew potential	8	5	0.1123825	128.40
Unconstrained dense coordinates followed by a fixed map to A				
Dense $z \mapsto A$, $m = 12$	12	9	0.0974009	121.50
Dense $z \mapsto A$, $m = 18$	18	15	0.0988612	117.91
Dense $z \mapsto A$, $m = 24$	24	21	0.1001312	127.38
Sum of independently predicted antisymmetric potentials				
$\sum_{s=1}^{2} A_s/\sqrt{2}$	12	9	0.1046972	126.65
$\sum_{s=1}^{3} A_s/\sqrt{3}$	18	15	0.1029831	115.75
$\sum_{s=1}^{4} A_s/2$	24	21	0.0986707	112.69
$\sum_{s=1}^{5} A_s/\sqrt{5}$	30	27	0.0986672	116.56
$\sum_{s=1}^{6} A_s/\sqrt{6}$	36	33	0.1047907	120.82
Unrestricted vector outer products, then antisymmetrized by the decoder				
$r = 2$ vector pairs	16	13	0.0951862	128.83
$r = 3$ vector pairs	24	21	0.0862193	101.95
$r = 4$ vector pairs	32	29	**0.0837907**	116.23
$r = 6$ vector pairs	48	45	0.0938667	115.80

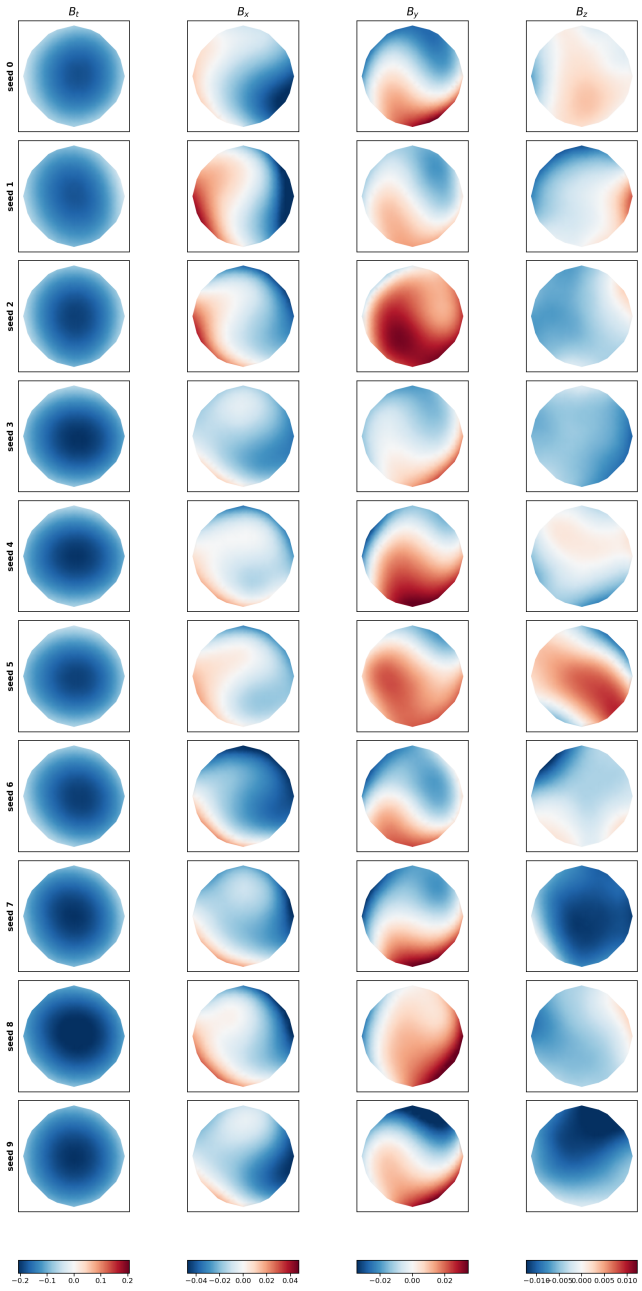
Table 7: Full controlled redundancy sweep from the original 10k-update Ball experiment. "Nullity" is the generic output-coordinate nullity of the map to the three physical current components, i.e., output coordinates minus rank 3. The deliberately overcomplete dense, multi-potential, and outer-product constructions mostly reach a lower loss than direct six-entry A at comparable training time.

K.3 ALL TEN INITIALIZATIONS: THREE REDUCED COMPONENTS VERSUS FOUR B AND SIX A COMPONENTS

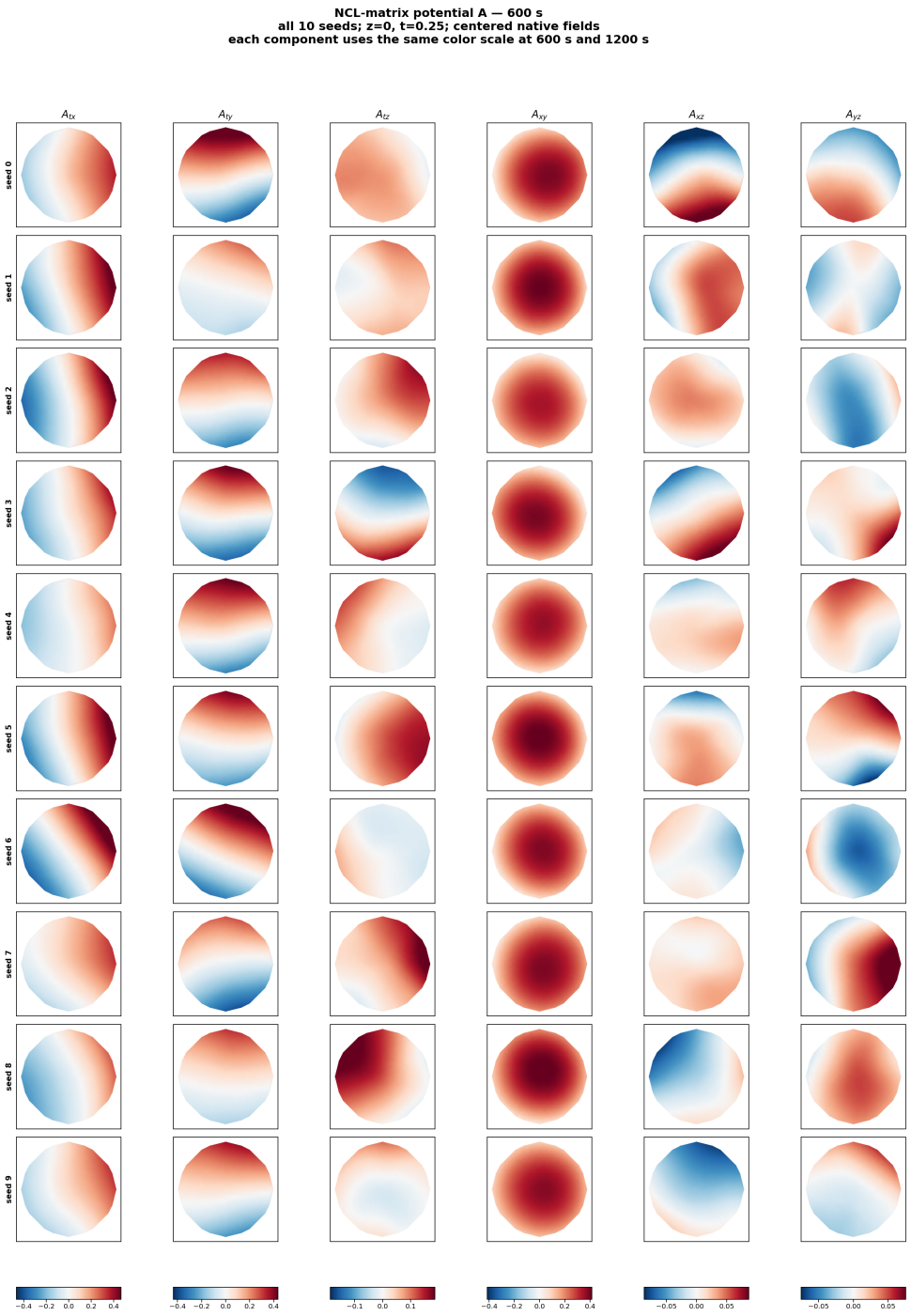
The following three displays show every learned native component at the same slice $z = 0$, $t = 0.25$, with one row per random initialization. They are the representation fields themselves, not the decoded physical current. The three reduced components learned by our method collapse to the same visible fields across the ten runs. In contrast, the four components of the NCL vector potential B and the six independent components of the antisymmetric NCL matrix potential A vary visibly from initialization to initialization. This is the direct visual distinction needed here: our 3-component coordinate has no gauge/null direction, whereas the 4-component B and 6-component A representations contain redundant degrees of freedom. Once decoded, potential differences that lie in those null directions do not alter the physical divergence-free field.



All ten learned three-component reduced fields. Each row is one independent initialization and the columns are the three intrinsic reduced components. At this common slice, the ten learned fields agree to visual resolution.



All ten learned four-component vector potentials B**.** Each row is one independent initialization and the columns are the four potential components. Unlike the intrinsic three-component coordinates, these native auxiliary fields visibly change between runs.



All ten learned six-component antisymmetric potentials A. Each row is one independent initialization and the columns are the six independent entries of A. The substantial run-to-run variation illustrates the redundant potential coordinates available before the fixed decoder maps them to the physical field.

K.4 WHY REDUNDANCY CAN HELP, AND WHY IT IS NOT A MONOTONIC RULE

At a nonzero Fourier mode, write a potential-based decoder abstractly as $\widehat{u}(k) = \mathcal{L}_k z(k)$, where z collects the learned auxiliary coordinates. If $\eta(k) \in \ker \mathcal{L}_k$, then

$$\mathcal{L}_k\big(z(k) + \eta(k)\big) = \mathcal{L}_k z(k) = \widehat{u}(k). \tag{271}$$

Consequently every physical optimum is an equivalence class $z^\star + \ker \mathcal{L}_k$, rather than a single auxiliary coordinate. In the $D = 4$ experiment, the minimal coordinates have three outputs, rank three, and no structural null direction. The vector-B construction has four outputs and one null direction, while the antisymmetric A construction has six independent components and three null directions. Optimization of a physical-field loss therefore takes place on a quotient: it can approach any member of the equivalent potential family instead of being required to select one particular A or B coordinate.

This mechanism gives a concrete explanation for a broad low-loss valley: the loss is constant to first order along exact gauge/null directions, and may be less sharply curved in nearby directions. Table 7 is the direct experimental check. Relative to direct six-entry A (0.1038496), the 12–24 coordinate dense constructions reach 0.0974009, 0.0988612, and 0.1001312; the unrestricted vector–outer-product construction reaches 0.0951862, 0.0862193, 0.0837907, and 0.0938667 for 2, 3, 4, and 6 vector pairs. The best record is therefore the 32-coordinate, 29-null-direction outer-product representation, which reaches 0.0837907 in 116.23 seconds, compared with 0.1038496 in 117.72 seconds for direct A.

The evidence nevertheless does not support a blanket monotonic law. Extra coordinates can also introduce cancellations, poor conditioning, or slow motion in weakly identified directions: the 48-coordinate outer-product version is still better than direct A but not better than the 32-coordinate version, and the multi-potential sweep has its best values at intermediate branch counts. The supported conclusion is that deliberately introduced redundancy creates many physically equivalent solutions and, in this controlled record, materially improves accessible loss and optimization speed over a useful range; the amount and decoder structure still matter.

K.5 TWO-DIMENSIONAL OPTIMAL TRANSPORT EVOLVING IN TIME

The stored optimal-transport diagnostic evolves a two-dimensional spatial transport field over time, so its conservation law is represented on the three-coordinate space–time domain (x, y, t). At 10k optimization steps, the original positive-density NCL branch obtained a loss of 5.989×10^{-1} in 374 seconds with $119{,}010$ parameters. The generic vector-B and matrix-A alternatives used $28{,}611$ parameters and obtained losses 4.931 and 8.613 in 296 and 287 seconds, respectively. This raw comparison is useful as a training diagnostic, but it is not a clean causal measurement of redundancy because the original NCL branch has a different positive-density construction and approximately four times as many parameters.

Within the generic overcomplete family itself, merely increasing coordinate count was again non-monotonic: the base dense representation attained a 10k loss of 8.396, whereas its 18- and 24-coordinate versions attained 64.50 and 78.88. The two experimental settings therefore agree on the appropriate interpretation: redundant coordinates can supply equivalent physical solutions and can improve optimization when the representation is well conditioned, but redundancy must be evaluated together with decoder structure, parameter budget, and numerical conditioning. The controlled Ball record above is the direct evidence for the landscape explanation; the optimal-transport record is retained as a complementary stress test with its differing density architecture stated explicitly.