BALANCENO: DISCOVERING WHAT TO CONSERVE FOR COUPLED PDE FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate autoregressive prediction of coupled partial differential equations is essential for fast multiphysics simulation, where spatial transport coexists with interactions that redistribute conserved quantities among fields. Existing surrogates face a critical dilemma: unconstrained regression suffers from severe physical drift over long rollouts, while hard-coded conservation laws are overly rigid, failing to capture the dynamic redistribution of quantities among coupled fields. We introduce BalanceNO, an operator neural framework that reformulates coupled-field evolution as a transport–interaction balance governed by a linear invariant subspace identified from trajectory data. BalanceNO discovers and validates exact balance constraints solely from training trajectories. To strictly enforce them during updates, it integrates telescoping spatial fluxes with nullspace-projected local interactions—preserving field-coupled invariants without freezing individual field totals. By embedding selective conservation directly into its core discrete update, BalanceNO bypasses the need for complete governing equations, seamlessly accommodating weighted conservation laws and partially conservative dynamics. Evaluated across complex spatiotemporal systems combining local transport, global nonlocal interactions, and strongly coupled fields in heterogeneous, three-dimensional, and interface-evolving media, BalanceNO delivers accurate long-horizon rollouts alongside roundoff-level invariant preservation. The code for the experiments can be found in https://anonymous.4open.science/r/BNO-0BB3.

1 INTRODUCTION

Learning accurate time-evolution operators for coupled partial differential equations (PDEs) is a central challenge in scientific machine learning, with applications ranging from reactive transport to multiphase, geophysical, and compressible flows. These systems evolve through both spatial transport and interactions among physical fields, making efficient surrogate prediction valuable wherever coupled simulations must be repeated. Their physical structure is not captured by treating every field as independently conserved. A successful surrogate must therefore predict spatial patterns and evolving component totals while respecting the combinations that remain invariant. During autoregressive deployment, this distinction must hold at each predicted step, rather than only approximately on training snapshots. The dominant learning paradigm nevertheless poses the task as a mapping from observed states or histories directly to subsequent solution fields.

In coupled dynamical systems, conservation often applies to combinations of field totals rather than to each field individually. Component totals may evolve through reaction, exchange, or momentum transfer while still satisfying global balance laws. This poses a challenge for autoregressive surrogates, which must distinguish spatial redistribution from admissible cross-field exchange and infer the conserved combinations from data. We therefore seek a discrete predictor that identifies these invariants from trajectories and enforces them exactly at every forecast step without suppressing valid component evolution. On a closed domain, applying a source-free conservative update to every field fixes every component integral and can exclude genuine exchange, reaction, or momentum evolution; adding unrestricted interactions removes that obstruction but can destroy the intended invariants.

We introduce BalanceNO to address this selective-conservation problem. Rather than learning an unrestricted state increment or imposing conservation independently on every output channel, Bal-

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.

anceNO learns a transport–interaction decomposition organized by a data-discovered invariant subspace. A scale-normalized subspace-identification procedure estimates linear combinations of spatially integrated fields that remain constant, with candidate constraints checked on held-out trajectories drawn exclusively from the training partition. The prediction operator then separates spatial transport from local interactions: shared face fluxes enter through a discrete divergence, while only the interaction term is projected into the nullspace of the identified conservation matrix. Flux cancellation preserves domain-integrated balances, and the projected interactions allow component totals to change in mutually compatible directions. Crucially, the projection does not constrain the entire state increment pointwise, so conserved densities can still move through the domain. Applying the discovered global invariants to local interactions is an explicit closure assumption rather than a consequence of trajectory conservation alone. Under this assumption, the core periodic-grid update preserves the identified invariant combinations algebraically, without using a PDE-residual penalty or requiring specification of the complete governing equations.

The technical distinction is not conservation enforcement alone: clawNO already accommodates mixed constrained and unconstrained outputs, and stoichiometric networks already represent balanced cross-species exchange. BalanceNO instead infers the conserved combinations and couples admissible interactions to a flux-based transport update, going beyond the source-free flux-learning setting studied by (U)NFV. The reported multi-seed comparisons use three two-dimensional coupled-PDE tasks—flow–phase-field coupling (Hohenberg & Halperin, 1977b; Cahn & Hilliard, 1958), rotating shallow water (Arakawa & Lamb, 1981; Sadourny, 1975), and heterogeneous reactive transport (Steefel et al., 2005; Ladd & Szymczak, 2021)—with compact, approximately parameter-matched implementations of ConvLSTM, U-Net, FNO, PINO, clawNO, LaMO, (U)NFV, and CoDA-NO. Separately, we instantiate the same BalanceNO construction on periodic compressible Navier–Stokes dynamics (Pirozzoli, 2011;?). In this conservative five-channel setting, the discovered invariant subspace is full rank, so $P_C = 0$ and BalanceNO reduces naturally to its conservative transport limit without changing the defining balance update. The three-dimensional realization therefore tests dimensional compatibility of the balance parameterization rather than the selective inter-field exchange emphasized by the two-dimensional benchmarks. One-step prediction assesses local approximation quality, autoregressive rollouts expose accumulated forecasting error, and invariant and field-total diagnostics distinguish correct conservation from artificially frozen component dynamics. Across the reported two-dimensional multi-seed comparisons, BalanceNO attains the lowest mean rollout error on each task while keeping invariant drift at roundoff level, demonstrating a favorable combination of predictive accuracy and selective conservation.

Main Contributions. We frame coupled-PDE forecasting around the distinction between conserved linear combinations and nonconserved component totals, rather than treating conservation as a channelwise property. We construct an operator whose training-only invariant identification, telescoping transport update, and nullspace-restricted interactions impose the resulting balances directly in discrete time. We provide a multi-seed evaluation across flow–phase-field coupling (Hohenberg & Halperin, 1977b; Cahn & Hilliard, 1958), rotating shallow water (Arakawa & Lamb, 1981; Sadourny, 1975), and heterogeneous reactive transport (Steefel et al., 2005; Ladd & Szymczak, 2021), establishing that the resulting model combines improved rollout accuracy with invariant preservation under the evaluated two-dimensional protocol. We additionally realize the same construction on periodic compressible Navier–Stokes dynamics (Pirozzoli, 2011;?), where the full-rank invariant subspace recovers the conservative limit without requiring a separate three-dimensional formulation.

2 Related Work

Expressive representations for coupled-field prediction. Convolutional predictors capture local spatial structure and temporal dependence through multiscale processing and recurrence, as exemplified by U-Net and ConvLSTM. Neural operators extend this representation capacity to function-to-function mappings: FNO parameterizes nonlocal interactions in Fourier space, LaMO combines latent representations with state-space processing, and CoDA-NO uses function-space attention across physical variables to support multiphysics learning and transfer. These designs improve the representation of spatial, temporal, and cross-variable dependencies without prescribing a particular governing equation. However, their standard prediction maps do not enforce a discrete balance identity, leaving conserved combinations vulnerable to drift during autoregressive prediction.

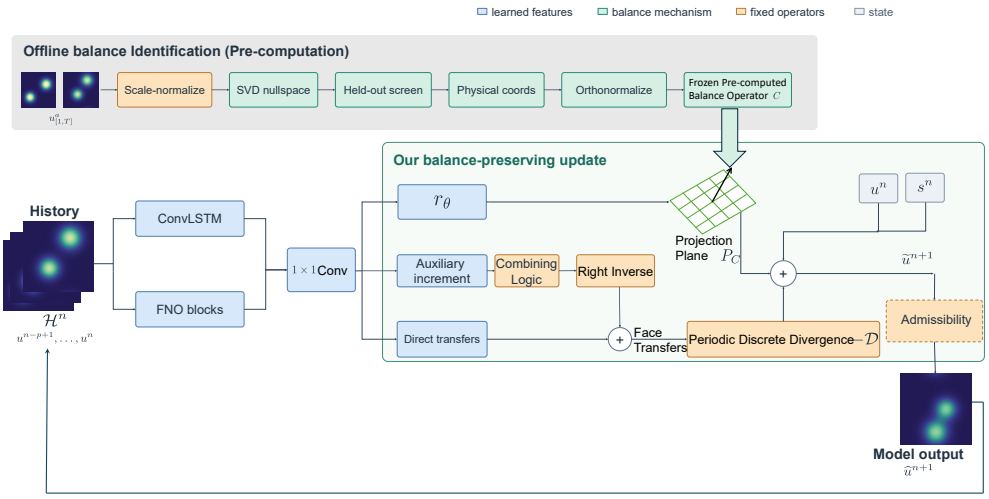

Figure 1: The pipeline of BalanceNO.

Physics-informed operator learning. Physics-informed methods supplement observations with differential relationships through the training objective. PINO combines operator learning with governing-equation residuals, including physics constraints evaluated at resolutions different from the observation grid. Such supervision can reduce dependence on labeled solutions and discourage predictions that violate the specified dynamics. However, the relevant differential relationships must be supplied, and finite-weight residual penalties do not by themselves guarantee exact discrete balance at inference.

Architectural conservation with prescribed structure. Hard-constrained methods restrict how a learned prediction can modify the physical state. FluxGNN incorporates conservative numerical principles into graph-based PDE solvers, while (U)NFV predicts numerical fluxes within a finite-volume update, using either supervised data or an unsupervised weak-form objective. ClawNO instead constructs divergence-free outputs through differential-form parameterizations to enforce continuity structure without requiring the complete governing equations. For reactive systems, fixed stoichiometric layers map learned interspecies transfers into balanced concentration changes. However, the conservation structure or stoichiometric relationships must be specified in advance, leaving the identification of unknown conserved field combinations as a separate modeling task.

3 METHOD

BalanceNO is constructed in four steps. We first state the desired global balance condition and derive a one-step update that satisfies it by construction. We then identify the unknown balance subspace from training trajectories. Given this subspace, a neural operator predicts the spatial-transfer and field-interaction coordinates used by the constrained update. Finally, we state the resulting structural guarantees and the learning procedure. This progression separates *what* must be preserved, *how* it is identified, and *how* it is enforced during prediction.

3.1 BALANCE-CONSTRAINED DISCRETE EVOLUTION

Let $u^n_{\boldsymbol{\xi}} \in \mathbb{R}^d$ denote the state of a coupled physical system at grid cell $\boldsymbol{\xi}$ and observation time t_n on a periodic Cartesian grid with G cells. Here $\boldsymbol{\xi}$ denotes a generic cell index, e.g., (i, j) in two dimensions and (i, j, k) in three dimensions. Given a history of p observations,

$$\mathcal{H}^n = (u^{n-p+1}, \ldots, u^n),$$

BalanceNO learns a discrete one-step flow map $\mathcal{G}_\theta : \mathcal{H}^n \mapsto \widehat{u}^{n+1}$. Here $\widehat{u}^{n+1}$ denotes the state returned by the full model and used in autoregressive rollout, whereas $\widetilde{u}^{n+1}$ below denotes the output

of the core balance update before any optional benchmark-specific fixed admissibility processing. When no such processing is used, $\widehat{u}^{n+1} = \widetilde{u}^{n+1}$. The formulation is intentionally discrete: the learned transfers represent one observation interval, and no continuous-time generator is assumed. During rollout, the predicted state is appended to the history and the oldest frame is discarded before applying the same one-step map again.

We consider linear integral quantities

$$m(u) = \sum_{\boldsymbol{\xi}} u_{\boldsymbol{\xi}}, \qquad \mathcal{I}_C(u) = C\, m(u), \qquad C \in \mathbb{R}^{K\times d}, \tag{1}$$

where the row space of C represents conserved combinations of field totals. A common cell-volume factor is omitted on the uniform grid. We therefore formulate the method in state coordinates in which the conserved quantities are linear combinations of field totals; spatially varying storage weights, when present, are incorporated into the state representation.

To derive the constrained update, consider the source-corrected internal increment $z^n = \widetilde{u}^{n+1} - u^n - s^n$, where s^n is an optional known external increment over one observation interval. We represent z^n as the sum of two admissible mechanisms. First, spatial transport is written as $-\mathcal{D}J_\theta$, where $\mathcal{D}$ is the periodic discrete divergence over all spatial directions and J_θ denotes the corresponding learned face transfers. Its spatial sum vanishes by telescoping, so it can redistribute each field without changing its global total. Second, local cross-field interaction is obtained by projecting an unrestricted learned increment r_θ onto $\ker C$. The corresponding orthogonal projector is $P_C = I_d - C^\top (CC^\top)^\dagger C$, where $(\cdot)^\dagger$ denotes the Moore–Penrose pseudoinverse. Combining these two mechanisms gives the defining BalanceNO update

$$\widetilde{u}^{n+1} = u^n - \mathcal{D}J_\theta[\mathcal{H}^n] + P_C r_\theta[\mathcal{H}^n] + s^n, \qquad P_C = I_d - C^\top (CC^\top)^\dagger C. \tag{2}$$

Thus, Equation equation 2 is a residual one-step update from u^n to $\widetilde{u}^{n+1}$: the neural operator determines the magnitudes and spatial patterns of transport and interaction, while the fixed operators determine which global changes are admissible. Cross-field dependence remains unrestricted because both J_θ and r_θ may depend on the complete coupled spatiotemporal history.

Equation equation 2 directly implies the balance identity stated later in Equation equation 6. Indeed, summing the update over the periodic grid and left-multiplying by C gives

$$\mathcal{I}_C(\widetilde{u}^{n+1}) - \mathcal{I}_C(u^n) - \mathcal{I}_C(s^n) = -C\sum_{\boldsymbol{\xi}} (\mathcal{D}J_\theta)_{\boldsymbol{\xi}} + CP_C \sum_{\boldsymbol{\xi}} (r_\theta)_{\boldsymbol{\xi}} = 0,$$

because $\sum_{\boldsymbol{\xi}} (\mathcal{D}J_\theta)_{\boldsymbol{\xi}} = 0$ for periodic boundaries and $CP_C = 0$. Equation equation 6 is therefore a consequence of the update in Equation equation 2, rather than a separate evolution rule.

Applying P_C pointwise defines a *local interaction closure*. The trajectories identify the global integral invariants, while the corresponding cellwise exchange law remains underdetermined. BalanceNO represents local non-transport interactions in the corresponding admissible field-space directions. This closure constrains the geometry of the update, not the receptive field of the neural operator. Proposition 2 below shows that, at the level of the combined state increment, this parameterization introduces no additional global restriction beyond the identified balance laws.

The remaining question is how the invariant subspace C is obtained when it is not known a priori.

3.2 Data-Driven Identification of the Balance Subspace

BalanceNO does not require the invariant coefficients C to be supplied. From training trajectories, we form source-corrected changes of spatially integrated states,

$$\delta m_{a,n} = m(u_a^{n+1}) - m(u_a^n) - b_a^n, \tag{3}$$

where a indexes trajectories and $b_a^n = m(s_a^n)$ when a known external increment is available; otherwise $b_a^n = 0$. A conserved row vector $c^\top$ must satisfy $c^\top \delta m_{a,n} \approx 0$ across transitions, so the desired row space of C is the approximate null space of these integrated changes.

To estimate this subspace, the fieldwise integrated changes are scale-normalized, a small-singular-value subspace is extracted on a discovery split, and that subspace is screened on disjoint held-out training trajectories. The accepted directions are mapped back to physical state coordinates

and orthonormalized to form C. Thus, the procedure identifies an invariant *subspace*; no physical meaning is assigned to a particular basis vector. If no direction survives screening, C is empty and $P_C = I_d$. The exact normalization, SVD screening rule, and numerical construction are given in Appendix A.1.

Invariant identification requires sufficient excitation of nonconserved integrated directions. Accordingly, candidate directions are screened on disjoint held-out training trajectories before C and the resulting projector P_C are fixed for neural-network training, and the architectural guarantee applies exactly to the identified subspace.

3.3 Neural Realization

With C fixed, the remaining learning problem is to predict expressive coordinates (J_θ, r_θ) for the admissible update in Equation equation 2. The balance parameterization itself is independent of a particular feature extractor. In the present implementation, a shared spatiotemporal operator combines convolutional recurrence with a Fourier neural-operator branch and produces a fused representation $F_\theta(\mathcal{H}^n)$. Learned heads decode this representation into spatial-transfer and interaction coordinates,

$$F_\theta(\mathcal{H}^n) \longmapsto \big(J_\theta, r_\theta\big), \tag{4}$$

which are not state predictions by themselves and are assembled only through Equation equation 2. The balance construction is independent of spatial dimension and of the particular feature extractor; the implementation uses the same recurrent–spectral design principle in two and three dimensions, with dimension-specific tensor operations and capacity choices deferred to the appendices.

For numerical conditioning, the implementation includes a fixed auxiliary path for conservative transport. A learned auxiliary increment is first made zero-mean; let a_0 denote the resulting increment. It is then mapped to an equivalent periodic face transfer by a fixed right inverse $\mathcal{R}$ of the negative discrete divergence, which satisfies $-\mathcal{D}(\mathcal{R}a_0) = a_0$ for every zero-mean increment a_0. For a general increment b, $\mathcal{R}$ acts after removal of the spatial mean $\langle b\rangle$, so that $-\mathcal{D}(\mathcal{R}b) = b - \langle b\rangle$; the mean component, which would change field totals, is never routed through this path. This defining property is shared across dimensions and is matched to the same discrete divergence $\mathcal{D}$ used in the balance update, whereas the concrete construction of $\mathcal{R}$ differs: prefix sums along the grid axes in two dimensions and Fourier-space inversion of the discrete Laplacian in three dimensions. In either case, $\mathcal{R}$ may be evaluated explicitly or through the equivalent identity above. In the three-dimensional realization, the complementary interaction component $(I_d - P_C)r_\theta$ is combined with the auxiliary coordinate before the zero-mean conservative transfer is formed. Its spatially varying contribution is thereby absorbed into the effective J_θ, while the admissible local interaction remains $P_C r_\theta$. This is only a reparameterization of the predictive coordinates and therefore leaves Equation equation 2, the admissible update space, and the balance identity unchanged. The concrete right-inverse constructions and decoder details are given in Appendix A.3; details of the recurrent and spectral branches and benchmark-specific fixed stabilization/admissibility operations are deferred to Appendices A.2–A.6.

The learned part of BalanceNO can therefore be summarized as

$$\mathcal{H}^n \xrightarrow{F_\theta} F \xrightarrow{\text{learned heads}} (J_\theta, r_\theta) \xrightarrow{-\mathcal{D}(\cdot)+P_C(\cdot)} \widetilde{u}^{n+1}. \tag{5}$$

Optional fixed admissibility processing, when required by a benchmark, is applied only after this balance update to obtain the model output $\widehat{u}^{n+1}$.

3.4 Structural Guarantees and Learning

The construction above yields two structural properties of the BalanceNO update. The first is the direct consequence of Equation equation 2 derived above.

Proposition 1 (Exact discrete balance). For periodic boundaries, the pre-admissibility state satisfies, in exact arithmetic,

$$\mathcal{I}_C(\widetilde{u}^{n+1}) = \mathcal{I}_C(u^n) + \mathcal{I}_C(s^n) \tag{6}$$

for every value of θ. Thus, Equation equation 6 is the invariant-level consequence of the one-step state update in Equation equation 2. A complete proof is provided in Appendix B. The constraint

is selective rather than fieldwise: individual field totals may change through $P_C r_\theta$ provided that the identified combinations remain invariant. This is essential for coupled systems in which internal interactions exchange or create individual state components while preserving only particular aggregate quantities.

Proposition 2 (Completeness of the balance parameterization). Let z be any discrete state increment satisfying $C \sum_{\boldsymbol{\xi}} z_{\boldsymbol{\xi}} = 0$. For unrestricted periodic face transfers and interaction increments, there exist J and r such that

$$z = -\mathcal{D}J + P_C r. \tag{7}$$

Hence the fixed balance layer introduces no additional global restriction beyond the identified conservation laws. The proof is given in Appendix B. The decomposition in Equation equation 7 is generally nonunique. Accordingly, J_θ and r_θ should be interpreted as structured predictive coordinates unless additional physical supervision or identifiability assumptions are introduced; BalanceNO guarantees the admissibility of the combined update, not unique recovery of physical fluxes or reaction rates.

The invariant matrix is identified once before neural-network optimization and is then fixed. Training uses only supervised state-prediction error over short rollouts; no conservation penalty and no governing-PDE residual are added to the BalanceNO objective. During autoregressive training and evaluation, the one-step map is applied repeatedly by shifting the history window and appending the latest prediction. Field normalization and detached pushforward rollout training are reported in Appendix A.5.

4 Experiments

We evaluate BalanceNO on coupled-field forecasting problems in which only selected linear combinations of field totals are conserved. On the two-dimensional benchmarks (Section 4.2), we examine whether BalanceNO produces accurate long-horizon rollouts, whether it preserves the identified invariants during autoregressive rollout, and whether selective conservation avoids the overconservation of fieldwise-conservative updates. We then test whether the same construction carries over to three dimensions (Section 4.3). Finally, a controlled ablation isolates the balance components from the backbone (Section 4.4).

4.1 Experimental Setup

Benchmarks. Three periodic two-dimensional systems exhibit selective conservation, each with a distinct difficulty for a learned surrogate: damped Navier–Stokes–Cahn–Hilliard dynamics with evolving material interfaces (NSCH) (Hohenberg & Halperin, 1977b; Cahn & Hilliard, 1958), rotating shallow-water flow with rotation-driven exchange between momentum components (SWE) (Arakawa & Lamb, 1981; Sadourny, 1975), and reaction–diffusion of $A + B \rightleftharpoons C$ in a heterogeneous porous medium (RD)(Steefel et al., 2005; Ladd & Szymczak, 2021). The conserved quantities are, respectively, the phase mass, the fluid mass, and the pair totals $q_A + q_C$ and $q_B + q_C$ of the stored species densities. The remaining totals evolve: the mean velocity decays under damping, the integrated momentum rotates under Coriolis forcing, and the individual species totals change through reaction. For the three-dimensional extension, we use periodic compressible Navier–Stokes dynamics (3D-CNS) (Pirozzoli, 2011;?). Governing equations, data generation, and splits are given in Appendices C.1 and C.2.

Baselines. We compare with U-Net (Ronneberger et al., 2015), ConvLSTM (Shi et al., 2015), FNO (Li et al., 2021), PINO (Li et al., 2024), LaMO (Tiwari et al., 2025), CoDA-NO (Rahman et al., 2024), a clawNO-style transport model (Liu et al., 2024), and an NFV-style neural finite-volume model inspired by (U)NFV (Lichtlé et al., 2025). PINO additionally minimizes a governing-equation residual. The clawNO-style and NFV-style models advance every field by a difference of face fluxes and therefore conserve each field total separately; we refer to them as *fieldwise-conservative controls*. All models are compact reimplementations trained from scratch under the same forecasting protocol with approximately matched parameter budgets (Appendix D.1).

Protocol and metrics. Every model observes two context frames and is rolled out autoregressively without access to future reference states. We report means and standard deviations over five seeds;

Table 1: Rollout relative L^2 errors on the two-dimensional benchmarks.

	NSCH		SWE		RD	
Model	macro	aggregate	macro	aggregate	macro	aggregate
BalanceNO	$\mathbf{8.8 \pm 4.8 \times 10^{-2}}$	$\mathbf{1.2 \pm 0.75 \times 10^{-1}}$	$\mathbf{2.2 \pm 0.15 \times 10^{-2}}$	$\mathbf{5.7 \pm 0.54 \times 10^{-3}}$	$\mathbf{1.1 \pm 0.12 \times 10^{-2}}$	$\mathbf{7.2 \pm 0.85 \times 10^{-3}}$
CoDA-NO	$5.7 \pm 0.71 \times 10^{-1}$	$9.6 \pm 0.71 \times 10^{-1}$	$6.4 \pm 0.079 \times 10^{-1}$	$1.6 \pm 0.099 \times 10^{-1}$	$9.3 \pm 1.4 \times 10^{-2}$	$6.3 \pm 0.94 \times 10^{-2}$
clawNO-style	1.1 ± 0.028	$8.5 \pm 0.12 \times 10^{-1}$	$7.7 \pm 0.15 \times 10^{-1}$	$2.0 \pm 0.075 \times 10^{-1}$	$2.9 \pm 0.12 \times 10^{-1}$	$1.7 \pm 0.070 \times 10^{-1}$
NFV-style	$6.3 \pm 0.62 \times 10^{1}$	$2.0 \pm 0.23 \times 10^{1}$	$8.0 \pm 0.15 \times 10^{-1}$	$2.0 \pm 0.11 \times 10^{-1}$	$2.6 \pm 0.11 \times 10^{-1}$	$1.6 \pm 0.064 \times 10^{-1}$
LaMO	$2.3 \pm 0.12 \times 10^{-1}$	$3.5 \pm 0.24 \times 10^{-1}$	$1.1 \pm 0.14 \times 10^{-1}$	$3.1 \pm 0.44 \times 10^{-2}$	$2.2 \pm 0.085 \times 10^{-2}$	$1.4 \pm 0.056 \times 10^{-2}$
FNO	$3.6 \pm 0.15 \times 10^{-1}$	$6.1 \pm 0.35 \times 10^{-1}$	$1.2 \pm 0.080 \times 10^{-1}$	$3.4 \pm 0.073 \times 10^{-2}$	$3.3 \pm 0.32 \times 10^{-2}$	$2.1 \pm 0.22 \times 10^{-2}$
PINO	$3.7 \pm 0.15 \times 10^{-1}$	$6.1 \pm 0.31 \times 10^{-1}$	$1.3 \pm 0.071 \times 10^{-1}$	$3.5 \pm 0.10 \times 10^{-2}$	$1.6 \pm 0.12 \times 10^{-1}$	$1.4 \pm 0.088 \times 10^{-1}$
U-Net	$3.8 \pm 0.56 \times 10^{-1}$	$5.4 \pm 0.60 \times 10^{-1}$	$2.2 \pm 0.13 \times 10^{-1}$	$5.8 \pm 0.50 \times 10^{-2}$	$4.0 \pm 0.51 \times 10^{-2}$	$2.5 \pm 0.34 \times 10^{-2}$
ConvLSTM	$4.4 \pm 0.24 \times 10^{-1}$	$6.3 \pm 0.25 \times 10^{-1}$	$2.9 \pm 0.14 \times 10^{-1}$	$8.0 \pm 0.44 \times 10^{-2}$	$2.2 \pm 0.10 \times 10^{-2}$	$1.4 \pm 0.084 \times 10^{-2}$

Table 2: Normalized drift of the conserved quantities over the rollout.

	SWE		RD		
Model	invariant	mass	invariant	q_A+q_C	q_B+q_C
BalanceNO	$\mathbf{2.9 \pm 0.70 \times 10^{-8}}$	$\mathbf{3.0 \pm 0.71 \times 10^{-8}}$	$\mathbf{4.6 \pm 0.80 \times 10^{-8}}$	$3.4 \pm 0.35 \times 10^{-8}$	$3.0 \pm 0.39 \times 10^{-8}$
CoDA-NO	$9.1 \pm 6.6 \times 10^{-3}$	$9.3 \pm 6.7 \times 10^{-3}$	$3.0 \pm 0.40 \times 10^{-2}$	$4.1 \pm 0.69 \times 10^{-2}$	$4.6 \pm 1.3 \times 10^{-2}$
clawNO-style	$\mathbf{2.9 \pm 0.81 \times 10^{-8}}$	$\mathbf{3.0 \pm 0.82 \times 10^{-8}}$	$\mathbf{4.4 \pm 0.58 \times 10^{-8}}$	$\mathbf{2.6 \pm 0.27 \times 10^{-8}}$	$\mathbf{2.7 \pm 0.26 \times 10^{-8}}$
NFV-style	$\mathbf{2.8 \pm 0.51 \times 10^{-8}}$	$\mathbf{2.9 \pm 0.52 \times 10^{-8}}$	$\mathbf{4.4 \pm 0.85 \times 10^{-8}}$	$\mathbf{2.7 \pm 0.49 \times 10^{-8}}$	$\mathbf{2.9 \pm 0.74 \times 10^{-8}}$
LaMO	$5.2 \pm 1.0 \times 10^{-3}$	$5.3 \pm 1.1 \times 10^{-3}$	$1.6 \pm 0.37 \times 10^{-3}$	$2.4 \pm 0.38 \times 10^{-3}$	$2.4 \pm 0.77 \times 10^{-3}$
FNO	$2.8 \pm 0.58 \times 10^{-3}$	$2.8 \pm 0.59 \times 10^{-3}$	$2.9 \pm 0.58 \times 10^{-3}$	$3.9 \pm 0.75 \times 10^{-3}$	$4.2 \pm 1.5 \times 10^{-3}$
PINO	$3.6 \pm 0.96 \times 10^{-3}$	$3.7 \pm 0.99 \times 10^{-3}$	$6.7 \pm 1.1 \times 10^{-2}$	$5.0 \pm 1.6 \times 10^{-2}$	$4.6 \pm 1.8 \times 10^{-2}$
U-Net	$5.9 \pm 1.1 \times 10^{-3}$	$5.9 \pm 1.2 \times 10^{-3}$	$2.8 \pm 1.5 \times 10^{-3}$	$4.0 \pm 1.7 \times 10^{-3}$	$4.1 \pm 3.2 \times 10^{-3}$
ConvLSTM	$2.3 \pm 0.42 \times 10^{-2}$	$2.3 \pm 0.42 \times 10^{-2}$	$3.6 \pm 0.51 \times 10^{-3}$	$6.4 \pm 0.28 \times 10^{-3}$	$5.0 \pm 1.3 \times 10^{-3}$

each seed changes both the generated data and the model initialization, and all models within a seed share the same splits. In all tables, the best mean in each column is bolded, and when other entries lie within $\pm 1\sigma$ of the best, they are bolded as well to indicate a statistical tie. Optimization hyperparameters are selected per benchmark by validation rollout macro relative L^2 error (Appendix D.3). Accuracy is measured by the rollout *macro* relative L^2 error, which averages channelwise relative errors, and the *aggregate* relative L^2 error, which takes one norm over all times, channels, and cells; the two can order models differently when channel scales differ. Balance is measured by two complementary diagnostics. *Invariant drift* is the departure of the discovered invariant combinations from their initial values, and *field-total error* is the error of every domain-integrated field against the reference. A model that freezes every field total attains roundoff-level invariant drift but a large field-total error whenever a nonconserved total evolves, so only the pair distinguishes correct selective conservation from overconservation. Definitions and additional metrics are given in Appendix D.4, and complete per-benchmark tables in Appendix E.

4.2 Evaluation on Two-Dimensional Benchmarks

Forecasting accuracy. Table 1 shows that BalanceNO attains the lowest mean macro and aggregate rollout error on all three benchmarks, reducing the macro error of the strongest baseline by roughly $2.6\times$ on NSCH, $5\times$ on SWE, and $2\times$ on RD. Seed variability is largest on NSCH, but the one-standard-deviation intervals of BalanceNO and of every baseline do not overlap.

Invariant preservation. Table 2 shows that invariant drift is determined by the structure of the update rather than by model capacity. BalanceNO and the two fieldwise-conservative controls remain near 10^{-8} over the full rollout, because their updates cancel the conserved combinations algebraically at every step. All direct predictors drift by 1.6×10^{-3} to 6.7×10^{-2}. This includes PINO, whose finite-weight residual penalty shapes training but leaves the output map free to violate the discrete balance. The guarantee covers only the identified linear invariants; it implies no incompressibility, energy, or free-energy property. Invariant drift alone also cannot distinguish BalanceNO from the fieldwise-conservative controls, which we examine next.

Selective versus fieldwise conservation. Fieldwise conservation is stronger than the balance structure of all three benchmarks. In SWE, the domain-integrated momentum (P_x, P_y) obeys

$$\dot{P}_x = fP_y, \qquad \dot{P}_y = -fP_x, \tag{8}$$

Table 3: Errors in domain-integrated quantities that evolve in the reference solution.

	field-total rel. error			SWE	
Model	NSCH	SWE	RD	momentum rel-L^2	energy rel. error
BalanceNO	$\mathbf{3.2 \pm 2.8 \times 10^{-2}}$	$\mathbf{8.3 \pm 0.91 \times 10^{-3}}$	$\mathbf{1.0 \pm 0.45 \times 10^{-3}}$	$\mathbf{9.2 \pm 1.4 \times 10^{-3}}$	$\mathbf{2.8 \pm 1.1 \times 10^{-4}}$
CoDA-NO	$1.1 \pm 0.40 \times 10^{-1}$	$8.2 \pm 0.45 \times 10^{-1}$	$6.3 \pm 0.98 \times 10^{-2}$	$8.7 \pm 0.21 \times 10^{-1}$	$1.8 \pm 1.3 \times 10^{-2}$
clawNO-style	$\mathbf{5.7 \pm 0.26 \times 10^{-2}}$	$8.1 \pm 0.55 \times 10^{-1}$	$2.5 \pm 0.13 \times 10^{-1}$	8.6×10^{-1}	$2.2 \pm 2.8 \times 10^{-2}$
NFV-style	$\mathbf{5.7 \pm 0.26 \times 10^{-2}}$	$8.1 \pm 0.55 \times 10^{-1}$	$2.5 \pm 0.13 \times 10^{-1}$	8.6×10^{-1}	$1.1 \pm 0.12 \times 10^{-2}$
LaMO	$\mathbf{4.4 \pm 0.83 \times 10^{-2}}$	$3.0 \pm 0.86 \times 10^{-2}$	$2.6 \pm 0.51 \times 10^{-3}$	$3.3 \pm 0.93 \times 10^{-2}$	$1.1 \pm 0.23 \times 10^{-2}$
FNO	$6.3 \pm 0.29 \times 10^{-2}$	$2.8 \pm 0.27 \times 10^{-2}$	$6.3 \pm 0.92 \times 10^{-3}$	$3.1 \pm 0.35 \times 10^{-2}$	$5.9 \pm 1.2 \times 10^{-3}$
PINO	$6.8 \pm 1.0 \times 10^{-2}$	$2.8 \pm 0.29 \times 10^{-2}$	$6.9 \pm 1.8 \times 10^{-2}$	$3.1 \pm 0.47 \times 10^{-2}$	$7.3 \pm 1.9 \times 10^{-3}$
U-Net	$6.1 \pm 1.4 \times 10^{-2}$	$1.1 \pm 0.25 \times 10^{-1}$	$1.5 \pm 0.59 \times 10^{-2}$	$1.2 \pm 0.15 \times 10^{-1}$	$1.3 \pm 0.28 \times 10^{-2}$
ConvLSTM	$1.7 \pm 0.20 \times 10^{-1}$	$1.3 \pm 0.085 \times 10^{-1}$	$5.6 \pm 0.65 \times 10^{-3}$	$1.5 \pm 0.067 \times 10^{-1}$	$4.7 \pm 0.85 \times 10^{-2}$

Table 4: Rollout errors and conservation on 3D compressible Navier–Stokes (64^3).

Model	macro rel-L^2	aggregate rel-L^2	final-step rel-L^2	field-total rel. error	invariant drift
BalanceNO	$\mathbf{1.0 \pm 0.096 \times 10^{-1}}$	$\mathbf{7.0 \pm 0.58 \times 10^{-2}}$	$\mathbf{1.3 \pm 0.19 \times 10^{-1}}$	$1.5 \pm 0.15 \times 10^{-8}$	$\mathbf{3.4 \pm 0.89 \times 10^{-8}}$
CoDA-NO	$4.7 \pm 0.13 \times 10^{-1}$	$3.7 \pm 0.054 \times 10^{-1}$	$5.0 \pm 0.057 \times 10^{-1}$	$1.1 \pm 0.46 \times 10^{-2}$	$1.4 \pm 0.58 \times 10^{-2}$
clawNO-style	$6.5 \pm 1.1 \times 10^{-1}$	$5.5 \pm 0.98 \times 10^{-1}$	1.4 ± 0.32	$\mathbf{1.3 \pm 0.10 \times 10^{-8}}$	$\mathbf{3.4 \pm 0.82 \times 10^{-8}}$
NFV-style	1.5 ± 0.20	1.9 ± 0.39	5.2 ± 1.3	$\mathbf{1.4 \pm 0.15 \times 10^{-8}}$	$\mathbf{3.4 \pm 0.76 \times 10^{-8}}$
LaMO	$2.2 \pm 0.041 \times 10^{-1}$	$1.8 \pm 0.032 \times 10^{-1}$	$3.0 \pm 0.045 \times 10^{-1}$	$1.7 \pm 0.097 \times 10^{-2}$	$1.4 \pm 0.23 \times 10^{-2}$
FNO	$3.2 \pm 0.076 \times 10^{-1}$	$2.4 \pm 0.060 \times 10^{-1}$	$3.4 \pm 0.10 \times 10^{-1}$	$2.8 \pm 0.26 \times 10^{-2}$	$2.3 \pm 0.76 \times 10^{-2}$
PINO	$3.1 \pm 0.12 \times 10^{-1}$	$2.4 \pm 0.091 \times 10^{-1}$	$3.4 \pm 0.084 \times 10^{-1}$	$2.5 \pm 0.19 \times 10^{-2}$	$2.1 \pm 0.28 \times 10^{-2}$
U-Net	$1.8 \pm 0.041 \times 10^{-1}$	$1.4 \pm 0.043 \times 10^{-1}$	$2.1 \pm 0.063 \times 10^{-1}$	$2.1 \pm 0.15 \times 10^{-2}$	$1.5 \pm 0.32 \times 10^{-2}$
ConvLSTM	$1.3 \pm 0.047 \times 10^{-1}$	$1.0 \pm 0.031 \times 10^{-1}$	$1.5 \pm 0.038 \times 10^{-1}$	$6.3 \pm 0.86 \times 10^{-3}$	$7.1 \pm 0.83 \times 10^{-3}$

so it rotates although mass is conserved. In RD, the reaction changes the stored species in the direction

$$(\Delta q_A, \Delta q_B, \Delta q_C) \propto (-1, -1, +1), \tag{9}$$

which preserves both pair totals while changing every species total. In NSCH, damping changes the mean velocity. A predictor that freezes every field total therefore incurs an irreducible field-total error whenever these totals evolve in the reference solution.

Table 3 confirms this prediction. The two fieldwise-conservative controls reach the same field-total error on every benchmark (to two significant digits) and identical integrated-momentum errors on SWE, which is the signature of their shared constraint rather than of comparable learned dynamics. BalanceNO preserves the same invariants at roundoff (Table 2) yet lowers the field-total error by two orders of magnitude on SWE and RD; on NSCH, the difference lies within one standard deviation. The reason is that $P_C r_\theta$ admits exactly the directions in which component totals may change: both the Coriolis rotation in Equation 8 and the reaction direction in Equation 9 lie in $\ker C$.

Conservation does not imply stability. On NSCH, the NFV-style control preserves the invariant, yet its rollout error grows to 6.3×10^1 (Table 1), because exact face-flux cancellation controls integrated quantities, not the amplification of nonconstant spatial modes. Together, Tables 2 and 3 show that near-exact invariant preservation can coexist with poor dynamics when an update is overconstrained, while an expressive unconstrained predictor can remain inaccurate in its integrated balances.

4.3 Extension to Three Dimensions

In 3D-CNS, mass, the three momentum components, and total energy are all conserved. The discovered invariant subspace is therefore full rank, $P_C = 0$, and BalanceNO reduces to its conservative transport limit without changing Equation 2. This setting tests the dimensional compatibility of the balance parameterization rather than selective exchange. Table 4 shows that BalanceNO attains the lowest rollout and final-step errors while keeping invariant drift at roundoff. Because every field total is invariant here, fieldwise conservation is no longer an overconstraint and the two controls also conserve at roundoff; their larger rollout errors therefore stem from their transport parameterizations rather than from overconservation.

Table 5: Controlled ablation on weighted exchange–advection–diffusion. Errors are percentages over a 40-step rollout; invariant drift is dimensionless. All variants have 6,816 parameters.

Variant	macro rel-L^2 (%)	aggregate rel-L^2 (%)	field-total rel. error (%)	invariant drift
BalanceNO (full)	$\mathbf{5.46}\ \pm\mathbf{4.27}$	$\mathbf{4.79}\ \pm\mathbf{3.78}$	$3.48\ \pm3.10$	$4.56\ \pm0.59 \times 10^{-9}$
No projection	$\mathbf{6.80}\ \pm\mathbf{5.64}$	$\mathbf{6.37}\ \pm\mathbf{5.57}$	$5.91\ \pm6.30$	$9.75\ \pm18.13 \times 10^{-3}$
Global projection	$\mathbf{6.16}\ \pm\mathbf{3.62}$	$\mathbf{6.23}\ \pm\mathbf{3.82}$	$3.95\ \pm2.55$	$3.78\ \pm0.54 \times 10^{-8}$
Divergence only	$61.02\ \pm4.67$	$45.70\ \pm3.42$	$128.81\ \pm11.98$	$\mathbf{3.39}\ \pm\mathbf{0.28} \times \mathbf{10^{-9}}$
Projection only	$23.70\ \pm2.29$	$26.77\ \pm3.03$	$\mathbf{1.38}\ \pm\mathbf{1.52}$	$\mathbf{3.44}\ \pm\mathbf{0.62} \times \mathbf{10^{-9}}$
No preconditioning	$18.69\ \pm12.68$	$15.79\ \pm11.08$	$3.81\ \pm2.46$	$4.12\ \pm0.58 \times 10^{-9}$

4.4 Ablation Study

To isolate the balance components from the backbone, we compare the full decoder with five variants that share the same backbone and learned heads and differ only in how the predicted coordinates form the state increment: *no projection*, *global projection* (a global mean correction), *divergence only*, *projection only*, and *no preconditioning* (without the auxiliary transfer path). The test system is a weighted exchange–advection–diffusion problem with the single invariant $\int (a + 2b)\,\mathrm{d}x$, in which every field total evolves and the conserved density is transported, so both fieldwise and pointwise conservation are incorrect constraints. Here the full decoder routes $(I_d - P_C)r_\theta$ into conservative transfers, as in the three-dimensional realization; the setup and its development history are detailed in Appendix F.

As Table 5 shows, the two incorrect constraints dominate the error despite roundoff-level conservation: divergence only reaches a 61.02% macro error against a 60.97% lower bound under fixed field totals, and projection only a 26.77% aggregate error against a 23.25% lower bound under a fixed conserved density. Removing the projection raises invariant drift to 9.8×10^{-3}, and removing the preconditioned path raises the macro error to 18.69%. The full decoder attains the lowest mean macro error, but no projection and global projection lie within one standard deviation, so the ablation supports selective constraints without establishing that local projection outperforms a global correction.

5 Conclusion

Discussion. Increasing spatial dimension does not alter the defining mathematical structure of BalanceNO. In every dimension, admissible source-corrected one-step increments are those satisfying the identified global balances; what changes is the geometry of the coordinates used to represent them. Higher-dimensional domains enlarge the space of possible flux coordinates while the divergence continues to map these coordinates to state increments, which can increase their nonuniqueness because multiple transfer fields may induce the same redistribution. At the same time, variations in the rank of the discovered invariant subspace change the relative degrees of freedom available to cross-field exchange and conservative redistribution. Consequently, transferring a particular coordinate parameterization across dimensions changes the redundancy and conditioning of the predictive coordinates even though the underlying balance constraints remain unchanged. Our extension addresses this distinction by preserving the same admissible update structure while reexpressing redundant or ineffective predictive coordinates through a fixed right inverse of the discrete divergence rather than removing them from the representable dynamics. Thus, the extension changes only how balance-compatible dynamics are parameterized. BalanceNO therefore retains a unified mathematical nature across dimensions: dimensionality affects the representation and conditioning of admissible dynamics, but not the underlying balance geometry, completeness of the constrained update space, or exact conservation principle.

Limitations and future work. Invariant discovery currently relies on a fixed SVD threshold and sufficiently informative trajectories; robustness under noisy, sparse, or weakly excited observations remains to be studied systematically. Future work could develop adaptive or uncertainty-aware subspace identification while retaining exact enforcement once the invariant subspace is determined.

AI USE STATEMENT

In this work, we used generative AI tools to polish the writing of the manuscript and to format tables for improved readability. We have reviewed all AI-assisted work: every AI-suggested edit to the text was checked by the authors for technical accuracy and consistency with our results, and every numerical value in the reformatted tables was verified against the original experimental records. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

Not applicable. This work develops surrogate models for the numerical simulation of coupled partial differential equations. All training and evaluation data are synthetic trajectories produced by numerical solvers; the work involves no human subjects, personal or sensitive data, or data collected from individuals, and we do not foresee direct harmful applications or conflicts of interest.

REPRODUCIBILITY STATEMENT

We have made the following efforts to ensure reproducibility. The BalanceNO update, the identification of the invariant subspace, and the neural realization are specified in Section 3, with the numerical invariant-discovery procedure in Appendix A.1. Complete proofs of the exact-balance and completeness results, together with the assumptions under which they hold, are given in Appendix B. Architectural details of the spatiotemporal backbone, the structured decoder and right-inverse constructions, the fixed admissibility operations, the training objective, and the selected hyperparameters are reported in Appendices A.2, A.3, A.4, A.5, and A.6, respectively. The governing equations and solver settings of all benchmarks are described in Appendix C.1, and the trajectory sampling and data splits in Appendix C.2. Baseline implementations, including the physics residual used by PINO, are described in Appendices D.1 and D.2; the optimization protocol and configuration selection in Appendix D.3; the evaluation metrics in Appendix D.4; and the hardware and software environment in Appendix D.5. Complete per-benchmark results over five seeds are reported in Appendix E. Source code for BalanceNO, all baselines, training, and evaluation is available at this anonymous repository; the construction of every dataset, including the reference solvers and seeded split generation, can be found in the code.

REFERENCES

Akio Arakawa and Vivian R Lamb. A potential enstrophy and energy conserving scheme for the shallow water equations. *Monthly Weather Review*, 109(1):18–36, 1981.

Johannes Brandstetter, Daniel E. Worrall, and Max Welling. Message passing neural PDE solvers. In *International Conference on Learning Representations (ICLR)*, 2022.

John W Cahn and John E Hilliard. Free energy of a nonuniform system. i. interfacial free energy. *The Journal of chemical physics*, 28(2):258–267, 1958.

Sigal Gottlieb, Chi-Wang Shu, and Eitan Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM Review*, 43:89–112, 2001. doi: 10.1137/S003614450036757X.

P. C. Hohenberg and B. I. Halperin. Theory of dynamic critical phenomena. *Reviews of Modern Physics*, 49:435–479, 1977a. doi: 10.1103/RevModPhys.49.435.

Pierre C Hohenberg and Bertrand I Halperin. Theory of dynamic critical phenomena. *Reviews of Modern Physics*, 49(3):435, 1977b.

Kevin Jamieson and Ameet Talwalkar. Non-stochastic best arm identification and hyperparameter optimization. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 240–248, 2016.

Anthony JC Ladd and Piotr Szymczak. Reactive flows in porous media: Challenges in theoretical and numerical methods. *Annual review of chemical and biomolecular engineering*, 12:543–571, 2021.

Randall J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, 2002. doi: 10.1017/CBO9780511791253.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations (ICLR)*, 2021.

Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM / IMS Journal of Data Science*, 2024. doi: 10.1145/3648506.

Nathan Lichtlé, Alexi Canesse, Zhe Fu, Hossein Nick Zinat Matin, Maria Laura Delle Monache, and Alexandre M. Bayen. (U)NFV: Supervised and unsupervised neural finite volume methods for solving hyperbolic PDEs. *arXiv preprint arXiv:2505.23702*, 2025.

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 *International Conference on Machine Learning (ICML)*, 2024.

Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. In *International Conference on Learning Representations (ICLR)*, 2017.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019.

Steven A. Orszag. On the elimination of aliasing in finite-difference schemes by filtering high-wavenumber components. *Journal of the Atmospheric Sciences*, 28:1074, 1971.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 8026–8037, 2019.

Sergio Pirozzoli. Numerical methods for high-speed flows. *Annual review of fluid mechanics*, 43 (1):163–194, 2011.

Md Ashiqur Rahman, Robert Joseph George, Mogab Elleithy, Daniel Leibovici, Zongyi Li, Boris Bonev, Colin White, Julius Berner, Raymond A. Yeh, Jean Kossaifi, Kamyar Azizzadenesheli, and Anima Anandkumar. Pretraining codomain attention neural operators for solving multiphysics PDEs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015. doi: 10.1007/978-3-319-24574-4_28.

Robert Sadourny. The dynamics of finite-difference models of the shallow-water equations. *Journal of the Atmospheric Sciences*, 32(4):680–689, 1975.

Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2015.

Carl I Steefel, Donald J DePaolo, and Peter C Lichtner. Reactive transport modeling: An essential tool and a new research approach for the earth sciences. *Earth and Planetary science letters*, 240 (3-4):539–558, 2005.

Karn Tiwari, Niladri Dutta, N M Anoop Krishnan, and A P Prathosh. Latent mamba operator for partial differential equations. *arXiv preprint arXiv:2505.19105*, 2025.

A IMPLEMENTATION DETAILS OF BALANCENO

This appendix specifies the components of BalanceNO introduced in Section 3: invariant-subspace discovery, the spatiotemporal backbone, the structured decoder, fixed admissibility operations, the training objective, and the selected settings.

A.1 INVARIANT-SUBSPACE DISCOVERY

This subsection gives the numerical procedure used to identify the linear invariant subspace described in Section 3. The same construction is used for two- and three-dimensional trajectories; only the set of spatial axes entering each field total changes. Let the source-corrected integrated-state changes from the discovery trajectories be stacked as rows of

$$D \in \mathbb{R}^{N_{\mathrm{dis}} \times d}.$$

To prevent fields with larger magnitudes from dominating the null-space calculation, we define

$$S = \mathrm{diag}(a_1, \ldots, a_d),$$

where a_c is the discovery-set average of the spatially integrated absolute magnitude of field c, lower-bounded by a small positive value. We then form

$$\widetilde{D} = DS^{-1}.$$

The normalized discovery matrix is decomposed by singular value decomposition,

$$\widetilde{D} = U\Sigma V^{\top}.$$

A right-singular direction is retained as a candidate invariant when

$$\frac{\sigma_j}{\sqrt{N_{\mathrm{dis}}}} \le \tau,$$

with $\tau = 10^{-5}$ in the reported experiments. If V_0 collects the accepted right-singular vectors, the candidate subspace is screened on a disjoint subset of training trajectories. Let $\widetilde{D}_{\mathrm{scr}}$ denote the normalized screening matrix. We compute

$$\widetilde{D}_{\mathrm{scr}} V_0^{\top} = U_{\mathrm{scr}} \Sigma_{\mathrm{scr}} W^{\top}$$

and retain only directions whose screening RMS singular value is at most 2τ. If W_0 denotes the retained screening directions, the invariant coefficients are mapped back to physical coordinates as

$$C = \mathrm{orth}_{\mathrm{rows}}\left(W_0 V_0 S^{-1}\right).$$

The implementation stores an orthonormal row basis, so the subsequent projector depends only on the discovered row space and not on the particular singular-vector basis.

If the available trajectories are insufficient, or if no candidate direction survives screening, no invariant is imposed and $P_C = I_d$. Discovery requires excitation of nonconserved integrated directions; nearly stationary trajectories can make some directions difficult to distinguish from invariants. A full-rank result $K = d$ is nevertheless valid when every represented field total is physically conserved, as in the unforced three-dimensional compressible Navier–Stokes setting. Screening reduces this ambiguity on the observed data, so the identified subspace is interpreted with respect to the sampled training ensemble.

For a nonempty C, the interaction projector is constructed in double precision as

$$P_C = I_d - C^{\top}(CC^{\top})^{\dagger}C,$$

explicitly symmetrized, converted to the model precision, and registered together with C as a non-trainable buffer.

A.2 SPATIOTEMPORAL BACKBONE

For the three two-dimensional benchmarks, BalanceNO uses a shared learned feature extractor. The input tensor has shape

$$X \in \mathbb{R}^{B \times p \times d \times H \times W}.$$

No coordinate or positional-encoding channels are appended.

Recurrent branch. Each history frame is lifted independently by the same circularly padded 3×3 convolution followed by GELU. The lifted sequence is processed by two stacked ConvLSTM cells. Both cells use circular 3×3 convolutions, hidden width w_ℓ, and zero initial hidden and memory states on every model invocation. Parameters are shared across history positions but not between the two recurrent layers. Only the final hidden state of the second cell, $h_p^{(2)}$, is retained.

Spectral branch. The history and field dimensions are flattened into pd channels and lifted by a 1×1 convolution to spectral width w_s. The representation is propagated through N_b residual Fourier blocks. For block ℓ,

$$Y_\ell = \mathrm{GN}_\ell(S_\ell), \qquad S_{\ell+1} = S_\ell + \mathrm{GELU}\left[\mathcal{F}^{-1}(\mathcal{W}_\ell \mathcal{F}(Y_\ell)) + \mathrm{Conv}_{1\times1,\ell}(Y_\ell)\right].$$

Two independent complex weight tensors act on the retained positive and negative bands of the first Fourier dimension and on the retained low-frequency band of the real-FFT dimension. Unretained spectral coefficients are set to zero, and the available modes are capped automatically by the grid resolution.

The recurrent and spectral representations are concatenated and fused by

$$F = \mathrm{Conv}_{1\times1,\, w_\ell + w_s \to w_\ell}\left(\left[h_p^{(2)}, S_{N_b}\right]\right),$$

without an additional activation. Spatial resolution is preserved throughout both branches.

Three-dimensional realization. For a three-dimensional volume the corresponding input is $X \in \mathbb{R}^{B\times p\times d\times N_z\times N_y\times N_x}$. The same recurrent–spectral organization is retained: two stacked periodic ConvLSTM cells are fused with a Fourier branch, with Conv2d operations replaced by Conv3d operations. The 3D Fourier convolution keeps the signed low-frequency corners in the two full FFT axes and the nonnegative low-frequency band in the real-FFT axis. In the selected native-64^3 profile, each recurrent input is augmented by its observed backward difference, scaled by a fixed factor of five; the first difference is zero. The latest lifted feature is also added to the second recurrent hidden state before fusion. These additions condition the learned coordinates on observed temporal changes and do not bypass the balance layer.

A.3 Structured Decoder and Transfer Preconditioning

In the two-dimensional realization, the fused feature map is decoded into balance coordinates rather than a direct state increment. A pointwise head predicts two raw face-transfer fields and an unrestricted interaction field, while a circular spatial head predicts an auxiliary transfer variable:

$$\begin{aligned}\left[J^{x,\mathrm{raw}}, J^{y,\mathrm{raw}}, r\right] &= \mathrm{Conv}_{1\times1,\, w_\ell \to 3d}\left[\mathrm{GELU}\left(\mathrm{Conv}_{1\times1,\, w_\ell\to w_\ell}(F)\right)\right], \\ a &= \mathrm{Conv}^{\mathrm{circ}}_{3\times3,\, w_\ell \to d}(F).\end{aligned}$$

Each output group has shape $B \times d \times H \times W$.

A discrete derivative suppresses slowly varying modes, which can make direct learning of low-frequency conservative increments poorly conditioned. The auxiliary path therefore reparameterizes such increments before the divergence is applied. The auxiliary output is low-pass filtered and stripped of its spatial mean,

$$a_0 = \mathcal{Z}\mathcal{B}_{k_{\mathrm{tr}}} a, \qquad \mathcal{Z}a = a - \langle a \rangle,$$

where $\mathcal{B}_{k_{\mathrm{tr}}}$ retains the same two-sided Fourier-band structure used by the spectral branch. The raw face transfers remain spectrally unrestricted.

A fixed periodic right inverse then maps a_0 to face transfers satisfying

$$-\mathcal{D}\mathcal{R}a_0 = a_0.$$

With

$$\rho_i = \frac{1}{W}\sum_j (a_0)_{ij},$$

the implementation uses

$$(J^{x,\mathrm{aux}})_{ij} = -\sum_{q=0}^{j} \big[(a_0)_{iq} - \rho_i\big], \qquad (J^{y,\mathrm{aux}})_{ij} = -\sum_{q=0}^{i} \rho_q.$$

The final face transfers are

$$J^x = J^{x,\mathrm{raw}} + J^{x,\mathrm{aux}}, \qquad J^y = J^{y,\mathrm{raw}} + J^{y,\mathrm{aux}}.$$

This auxiliary construction does not add a new physical constraint; it provides an explicit coordinate path for zero-mean conservative increments.

For periodic boundaries, the discrete divergence is

$$(\mathcal{D}J)_{ij} = J^x_{ij} - J^x_{i,j-1} + J^y_{ij} - J^y_{i-1,j},$$

implemented by circular shifts. Its spatial sum vanishes identically. The interaction head is transformed pointwise as

$$r'_{ij} = P_C r_{ij}.$$

The latest input frame provides the residual state in the update. A supplied source, when present, is added in state-increment units immediately before optional admissibility processing.

Three-dimensional realization. The 3D decoder follows the same balance decomposition with three face directions. A pointwise global head predicts $(J^{x,\mathrm{raw}}, J^{y,\mathrm{raw}}, J^{z,\mathrm{raw}}, r)$ and a periodic $3 \times 3 \times 3$ head predicts the auxiliary increment a. The selected native-64^3 profile additionally uses one local face network per direction. Each local network sees the latest state and observed history difference together with the fused features, and evaluates a $2 \times 3 \times 3$ stencil oriented normal to that face followed by pointwise layers. These local heads only augment the raw face transfers and therefore remain inside the conservative divergence.

The unrestricted interaction is split into $P_C r$ and its complementary component $(I_d - P_C)r$. The former is retained as the local interaction in Equation equation 2; the latter is combined with the auxiliary coordinate before conservative transfer construction,

$$b = \mathcal{B}_{k_{\mathrm{tr}}}\big(a + (I_d - P_C)r\big), \qquad J = J^{\mathrm{raw}} + J^{\mathrm{local}} + \mathcal{R}b,$$

where $\mathcal{B}_{k_{\mathrm{tr}}}$ is the optional three-dimensional low-pass filter. The right inverse is defined after removal of the DC component, so $-\mathcal{D}\mathcal{R}b = b - \langle b\rangle$. For a discrete Fourier mode $\boldsymbol{k} = (k_x, k_y, k_z) \neq 0$, the explicit implementation uses

$$\lambda(\boldsymbol{k}) = -\sum_{q\in\{x,y,z\}} 4\sin^2(k_q/2), \qquad \widehat{(\mathcal{R}b)}_q(\boldsymbol{k}) = \frac{1 - e^{\mathrm{i}k_q}}{\lambda(\boldsymbol{k})}\,\widehat{b}(\boldsymbol{k}),$$

with zero DC mode. The symbols match the same backward-difference divergence used in the update, including nonzero Nyquist face gradients. The production path may evaluate the algebraically identical relation $-\mathcal{D}\mathcal{R}b = b - \langle b\rangle$ directly, avoiding explicit construction of the auxiliary face field. Raw and local face-transfer heads remain spectrally unrestricted. Consequently, routing $(I_d - P_C)r$ through this path changes only the coordinate parameterization of J_θ and not the balance law. In the full-rank case $P_C = 0$, as expected for the unforced five-channel compressible Navier–Stokes data, the full and divergence-only balance routes are therefore identical for shared weights.

A.4 Admissibility and Fixed Stabilization

For the two-dimensional benchmarks containing a state variable that must remain positive, BalanceNO optionally applies a fixed mean-preserving rescaling after the balance update. For a designated field c, let μ_c and ν_c denote its spatial mean and minimum in the pre-admissibility state v. With positivity floor ϵ_p and numerical stabilizer ϵ_d,

$$\mu_c^+ = \max(\mu_c, \epsilon_p),$$
$$\eta_c = \mathrm{clip}_{[0,1]}\left(\frac{\mu_c^+ - \epsilon_p}{\mu_c^+ - \nu_c + \epsilon_d}\right),$$
$$\mathcal{L}(v)_c = \begin{cases} \mu_c^+ + \eta_c(v_c - \mu_c), & \nu_c < \epsilon_p, \\ v_c, & \text{otherwise.} \end{cases}$$

The implementation uses $\epsilon_p = 10^{-6}$ and $\epsilon_d = 10^{-12}$. When $\mu_c \geq \epsilon_p$, the operation preserves the field mean and therefore preserves any balance involving that field total. If the pre-limited mean itself lies below the floor, enforcing the positivity floor necessarily changes that field total.

The rotating shallow-water benchmark additionally uses a fixed periodic five-point Laplacian,

$$\alpha \Delta_h u^n,$$

in the state update. Its spatial sum is zero. For the dimensionless five-point stencil, the isolated map $I + \alpha\Delta_h$ is nonexpansive in the discrete Euclidean norm for $0 \leq \alpha \leq 1/4$; the implementation restricts α to $[0, 0.24]$. This nonexpansiveness statement applies to the isolated smoothing map. Likewise, BalanceNO's structural guarantee is restricted to the identified integral invariants; pointwise incompressibility, entropy inequalities, and nonlinear energy conservation are separate PDE-specific properties.

A.5 Training Objective

The invariant matrix is identified before neural-network optimization and remains fixed throughout training. Let ω_c denote the training-set standard deviation of one-step temporal increments of field c, lower-bounded by $\omega_{\min} = 10^{-4}$ in the implementation. The normalized state error is

$$\ell(\widehat{u}, u) = \frac{1}{Gd} \sum_{\boldsymbol{\xi}, c} \left(\frac{\widehat{u}_{\boldsymbol{\xi} c} - u_{\boldsymbol{\xi} c}}{\omega_c} \right)^2 .$$

For a rollout horizon R, the training objective is

$$\mathcal{J}(\theta) = \mathbb{E} \left[\frac{1}{R} \sum_{h=0}^{R-1} (1 + 0.25h)\, \ell \left(\mathcal{G}_\theta(\mathcal{H}_h), u^{n+h+1} \right) \right].$$

After each forecast, the oldest state is removed and the new prediction is appended to the history. For subsequent rollout steps, the appended forecast is detached from the computational graph,

$$\mathcal{H}_{h+1} = \text{update} \left(\mathcal{H}_h, \text{sg}(\widehat{u}^{n+h+1}) \right),$$

where sg is the identity in the forward pass and has zero derivative. Thus later rollout losses expose the model to its own forecast distribution without backpropagating through earlier forecast transitions.

BalanceNO uses no conservation penalty and no governing-PDE residual. Benchmark-specific optimization schedules are applied uniformly to all compared models within each benchmark.

For the memory-bounded native-64^3 continuation, the same objective is estimated from uniformly sampled spatial output sites. BalanceNO still computes its global recurrent–spectral context, Fourier filtering, and spatial mean on the complete periodic volume; local decoder heads are evaluated only on the sampled region and the required true periodic halo. Complete predicted volumes, detached between rollout steps as above, form the next autoregressive history. Thus spatial sampling changes the supervised estimator and its variance, not the balance operator or the domain on which conservation is enforced.

A.6 Architecture and Implementation Settings

BalanceNO uses the same learned architecture for all three two-dimensional benchmark systems. The shared two-dimensional configuration is

$$p = 2, \qquad w_\ell = 14, \qquad w_s = 9, \qquad N_b = 2, \qquad k_{\max} = 7,$$

where p is the input-history length, w_ℓ and w_s are the recurrent and spectral widths, respectively, N_b is the number of residual Fourier blocks, and $k_{\max}$ is the number of retained Fourier modes.

Benchmark-dependent settings occur only in fixed decoder or admissibility operations. For Navier–Stokes–Cahn–Hilliard, $(k_{\text{tr}}, \alpha) = (6, 0)$ and no positivity limiter is used. For rotating shallow water, $(k_{\text{tr}}, \alpha) = (7, 0.12)$ and the height field h is positivity constrained. For heterogeneous porous

reaction–diffusion, $(k_{\mathrm{tr}}, \alpha) = (7, 0)$ and the three species-storage channels (q_A, q_B, q_C) are positivity constrained, while the static porosity channel is not. These choices do not alter the learned encoder, spectral operator, fusion module, or output-head dimensions.

The two-dimensional protocol uses 64×64 grids, with 64 training, 16 validation, and 16 test trajectories per seed. The porous dataset contains 12 frames per trajectory, the fluid–phase-field dataset contains 48, and the shallow-water dataset contains 24.

Selected three-dimensional BalanceNO profile. The supplied native-64^3 configuration keeps $p = 2$, uses four retained Fourier modes, two residual Fourier blocks, and an auxiliary transfer cutoff $k_{\mathrm{tr}} = 4$. Under the common target of $250{,}000$ real trainable parameters, the 3D factory selects recurrent width $w_\ell = 8$, spectral width $w_s = 5$, and local face-head width 85, giving $250{,}904$ trainable real scalars. The recurrent gates use periodic $3 \times 3 \times 3$ kernels, the global coordinate head has two pointwise stages, and each local face head has depth six. The selected switches enable the local latest-frame skip, difference-augmented lifting, the zero-initialized temporal interaction shortcut, local face-history coordinates, fused-feature conditioning of the local face heads, and the exact fused evaluation of the right-inverse identity. The observed history differences used by the lift and local face coordinates are scaled by five, while the zero-initialized temporal interaction shortcut receives the fixed feature $10(u^n - u^{n-1})$. No coarse feature grid, scalar positivity limiter, or fixed artificial diffusion is used. These choices alter capacity allocation and predictive coordinates only; the defining update remains Equation equation 2.

Output-head initialization. The final convolution in each balance-coordinate output head is initialized with zero bias and Gaussian weights with standard deviation 10^{-3}; the 3D local face-head readouts use the same initialization.

Right-inverse evaluation. In two dimensions, reductions and cumulative sums in the fixed right-inverse construction are performed independently for each sample and field. The vertically accumulated row mean is broadcast across the horizontal axis, matching the implementation of the periodic transfer construction. In three dimensions, the explicit right inverse uses the matching discrete Fourier symbols given in Appendix A.3; the selected production path evaluates the exact identity $-\mathcal{D}\mathcal{R}b = b - \langle b \rangle$ directly.

B Proofs and Scope of the Balance Guarantees

Proof of Proposition 1. Summing the periodic divergence over the grid gives

$$\sum_{\boldsymbol{\xi}} (\mathcal{D}J)_{\boldsymbol{\xi}} = 0.$$

For the projected interaction,

$$C\, m(P_C r) = C P_C\, m(r) = 0.$$

Therefore the core BalanceNO update satisfies

$$\mathcal{I}_C(\widetilde{u}^{n+1}) = \mathcal{I}_C(u^n) + \mathcal{I}_C(s^n),$$

independently of the learned network parameters. □

Proof of Proposition 2. Let z satisfy

$$C \sum_{\boldsymbol{\xi}} z_{\boldsymbol{\xi}} = 0$$

and define its spatial mean

$$\bar{z} = \frac{1}{G} \sum_{\boldsymbol{\xi}} z_{\boldsymbol{\xi}}.$$

Then $C\bar{z} = 0$, so $P_C \bar{z} = \bar{z}$. The residual $z - \bar{z}$ has zero spatial mean and therefore lies in the range of the periodic discrete divergence. Choosing $r_{\boldsymbol{\xi}} = \bar{z}$ at every grid cell and any periodic right-inverse flux J satisfying

$$-\mathcal{D}J = z - \bar{z}$$

gives

$$z = -\mathcal{D}J + P_C r.$$

Thus the balance layer spans all discrete increments compatible with the identified invariant subspace. □

Proposition 2 characterizes the range of the fixed parameterization; approximation by a finite neural network is a separate question. The transfer–interaction decomposition is generally nonunique, so the learned coordinates are interpreted as predictive coordinates rather than as unique physical fluxes or reaction rates without additional assumptions or supervision.

The architectural guarantee is exact for the identified invariant subspace. If $C_\star$ denotes a target invariant matrix and $P_\star$ its nullspace projector, the defect relative to that target under an estimated projector P_C satisfies

$$\left\| C_\star \left[m(\widetilde{u}^{n+1}) - m(u^n) - m(s^n) \right] \right\|_2 \le \|C_\star\|_2 \, \|P_C - P_\star\|_2 \, \|m(r_\theta)\|_2.$$

Thus, sensitivity to invariant-subspace estimation is quantified explicitly by the projector mismatch $\|P_C - P_\star\|_2$.

C Benchmark Details

C.1 Benchmark Systems and Reference Data

The three two-dimensional benchmarks are periodic systems posed on the unit square $[0, 1)^2$ and discretized on a uniform 64×64 grid. The supplied three-dimensional extension additionally considers unforced compressible Navier–Stokes dynamics on the periodic unit cube $[0, 1)^3$. Each system is written in the storage variables that the surrogate actually observes, so that the quantities conserved by the reference dynamics are *linear* in the supplied channels, as Equation equation 1 requires. For the three two-dimensional systems, a single dissipation scale $\nu_0 = 10^{-2}$ sets the viscosity and diffusivity magnitudes. Table 6 collects the two-dimensional discretization and horizon settings; the three-dimensional generator is described separately below. To keep the notation unambiguous, we write $\varepsilon_{\mathrm{int}}$ for the Cahn–Hilliard interface width and $\theta(x)$ for the porosity field. The symbols C, ϕ, and ϵ_p retain the meanings assigned in Section 3 and Appendix A.4.

Table 6: Reference-data settings for the three two-dimensional benchmarks. d counts state channels, of which the porous system has one static channel. The rollout length is the number of autoregressive steps evaluated after the $p = 2$ observed context frames.

System	d	Conserved integrals	Δt	Frames	Rollout
Navier–Stokes–Cahn–Hilliard	3	$\sum_{ij} \phi$	0.04	48	46
Rotating shallow water	3	$\sum_{ij} h$	0.04	24	22
Heterogeneous porous reaction–diffusion	$3 + 1$	$\sum_{ij}(q_A + q_C), \sum_{ij}(q_B + q_C), \sum_{ij} \theta$	0.10	12	10

Shared numerical settings. For the three two-dimensional benchmarks, every observation interval Δt is advanced with five solver substeps, $\Delta t_{\mathrm{sub}} = \Delta t/5$. The shallow-water and reaction–diffusion generators additionally subdivide each substep against an explicit stability bound, so their effective step is $\min(\Delta t_{\mathrm{rem}}, \Delta t_{\mathrm{sub}}, \Delta t_{\mathrm{stab}})$. The spectral solvers apply the two-thirds box dealiasing mask $|k_x|, |k_y| \le N/3$ to every nonlinear product (Orszag, 1971). Integration is carried out in double precision and trajectories are stored in single precision. Each generator aborts if the state becomes non-finite or leaves the admissible set (nonpositive height or species storage), so no clipping is applied that would destroy a conserved integral.

Random initial states. For the three two-dimensional benchmarks, all initial conditions are built from smooth Gaussian random fields. Independent standard normal noise is low-pass filtered in Fourier space by $\exp[-\frac{1}{2}(|k|^2/k_{\mathrm{c}}^2)^2]$ and then shifted and rescaled to zero mean and unit variance per sample, yielding fields $s_0, s_1, \ldots$; the cutoff k_{c} is specified per benchmark below. Drawing the domain means separately from the fluctuation pattern is deliberate: the invariant-discovery stage of Appendix A.1 requires the nonconserved integrated directions to be excited, which a fixed-mean ensemble would not provide.

Navier–Stokes–Cahn–Hilliard dynamics. A matched-density two-phase flow of Model H type (Hohenberg & Halperin, 1977a), built on the Cahn–Hilliard free energy of Cahn & Hilliard (1958), couples a damped incompressible velocity to a conserved phase field,

$$\begin{aligned} &\partial_t \mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u} = -\nabla p + \nu\Delta\mathbf{u} - \lambda\mathbf{u} - \sigma\phi\nabla\mu, \qquad \nabla\cdot\mathbf{u} = 0, \\ &\partial_t \phi + \mathbf{u}\cdot\nabla\phi = M\Delta\mu, \qquad \mu = \phi^3 - \phi - \varepsilon_{\mathrm{int}}^2\Delta\phi, \end{aligned} \tag{10}$$

with mobility $M = 0.02$, interface width $\varepsilon_{\mathrm{int}} = 0.04$, capillary coefficient $\sigma = 0.05$, drag $\lambda = 0.10$, and viscosity $\nu = \nu_0$. The observed fields are (u, v, ϕ). The phase mass $\sum_{ij}\phi$ is conserved, whereas the linear drag makes the two velocity integrals decay; this is the asymmetry that a fieldwise-conservative update cannot represent. Initial conditions use $k_{\mathrm{c}} = 3.0$. The phase field is $\phi = 0.36\tanh s_0$ shifted to a prescribed mean $\sim \mathcal{U}(-0.12, 0.12)$. The velocity is the skew gradient $(\partial_y\psi, -\partial_x\psi)$ of a random stream function $\psi = s_1$, rescaled to root-mean-square speed $\sim \mathcal{U}(0.18, 0.32)$, after which an independent mean drift $\sim \mathcal{U}(-0.12, 0.12)$ is added to each component. The added drift is what excites the damped mean-velocity directions. Each substep is a dealiased Fourier semi-implicit update: advection and the mean-removed capillary force $-\sigma\phi\nabla\mu$ are treated explicitly, the viscous and drag terms are absorbed into the implicit factor $1+\Delta t(\nu k^2+\lambda)$, the velocity is made divergence-free by a Leray projection, and the biharmonic part of the phase equation is absorbed into $1 + \Delta t\, M\varepsilon_{\mathrm{int}}^2 k^4$. The $k = 0$ coefficient of ϕ is carried through each substep unchanged, so $\sum_{ij}\phi$ is exactly constant along the reference trajectory.

Rotating shallow-water flow. The conservative shallow-water system on a constant-f plane evolves (h, hu, hv) under

$$\partial_t \begin{pmatrix} h \\ hu \\ hv \end{pmatrix} + \nabla\cdot \begin{pmatrix} hu & hv \\ hu^2 + \frac{1}{2}gh^2 & huv \\ huv & hv^2 + \frac{1}{2}gh^2 \end{pmatrix} = \begin{pmatrix} 0 \\ f\,hv \\ -f\,hu \end{pmatrix}, \tag{11}$$

with $g = 1$ and $f = 2$. Mass $\sum_{ij} h$ is conserved, while the Coriolis source rotates the integrated momentum according to Equation equation 8. Initial conditions use $k_{\mathrm{c}} = 3.0$ with $\bar{h} \sim \mathcal{U}(0.9, 1.1)$ and $h = \bar{h}\exp(0.07\, s_0)$ rescaled to mean $\bar{h}$. A domain-mean velocity of magnitude $\sim \mathcal{U}(0.08, 0.24)$ and uniformly random heading $\sim \mathcal{U}(0, 2\pi)$ is superposed with fluctuations $0.08\, s_1$ and $0.08\, s_2$, and the momenta are stored as hu and hv. The nonzero mean momentum and random heading ensure that the rotation of Equation equation 8 is visible in every trajectory rather than only in the fluctuation field. Each step is a second-order Strang composition: an exact half-step rotation of (hu, hv) through angle $f\Delta t/2$, an SSPRK2 transport step (Gottlieb et al., 2001), and a second exact half-step rotation. The spatial operator is a periodic local Lax–Friedrichs (Rusanov) finite-volume divergence (LeVeque, 2002) on $\Delta x = 1/64$, with face dissipation set by the larger of $|u| + \sqrt{gh}$ on the two cells adjoining each face. The step is limited by $\Delta t_{\mathrm{stab}} = \mathrm{CFL}/\max_{ij}[(|u| + |v| + 2\sqrt{gh})/\Delta x]$ with $\mathrm{CFL} = 0.35$. Because the flux form telescopes and the rotation half-steps act only on momentum, mass is conserved to roundoff without any height clipping.

Heterogeneous porous reaction–diffusion. A reversible reaction $A + B \rightleftharpoons C$ proceeds in a medium of spatially varying porosity,

$$\theta(x)\,\partial_t c_i = \nabla\cdot(\theta(x) D_i^0 \nabla c_i) + \theta(x) R_i(c), \qquad R = (-r, -r, +r), \qquad r = k_f c_A c_B - k_r c_C, \tag{12}$$

with base diffusivities $D^0 = \nu_0(1.0, 0.6, 0.35)$, forward rate $k_f = 4.0$, and reverse rate $k_r = 1.0$. The porosity is $\theta = \theta_0 + \theta_1\tanh(1.6\, s)$ with $\theta_0 \sim \mathcal{U}(0.55, 0.72)$, $\theta_1 \sim \mathcal{U}(0.30, 0.45)$ and $k_{\mathrm{c}} = 3.0$, clipped to $[0.2, 1.0]$; it is static in time and supplied to every model as a fourth input channel. Concentrations are initialized with $k_{\mathrm{c}} = 3.5$ and means $\bar{c} = (\mathcal{U}(0.35, 0.60), \mathcal{U}(0.35, 0.60), \mathcal{U}(0.05, 0.20))$, as $\bar{c}_i \exp[\mathrm{clip}(0.35\, s_i, -0.7, 0.7)]$ rescaled to mean $\bar{c}_i$. The deliberately low initial abundance of C keeps the forward reaction active over the recorded horizon. The generator integrates Equation equation 12 in the better-conditioned concentration variable with SSPRK2 (Gottlieb et al., 2001) and the adaptive bound

$$\Delta t_{\mathrm{stab}} = \frac{0.45}{D_{\max}^0 k_{\max}^2 + 2D_{\max}^0 k_{\max} + k_f c_{\max} + k_r},$$

where the factor $2k_{\max}$ conservatively bounds the variable-coefficient term $|\nabla\theta/\theta|$ in grid-wavenumber units. Trajectories are stored in the molar-density variables $q_i = \theta c_i$. This choice is

what makes the benchmark expressible in the present framework: because θ is static, $\partial_t q_i = \theta\, \partial_t c_i$, the dealiased spectral divergence has exactly zero spatial mean, and the reaction vector cancels pointwise on the combinations $A + C$ and $B + C$. The two physical reaction invariants $\sum_{ij}(q_A + q_C)$ and $\sum_{ij}(q_B + q_C)$ are therefore linear in the supplied channels, whereas the corresponding combinations of the concentrations c_i are not. Because θ is supplied as a fourth, pointwise-static state channel, the discovery procedure also identifies its total $\sum_{ij}\theta$ as a third invariant direction. This is a concrete instance of the requirement stated below Equation equation 1 that spatially varying storage weights be absorbed into the state representation.

Three-dimensional compressible Navier–Stokes. The three-dimensional extension uses the conservative variables

$$u = (\rho, \rho u, \rho v, \rho w, E), \qquad E = \frac{p}{\gamma - 1} + \frac{\rho}{2}(u^2 + v^2 + w^2),$$

on the periodic unit cube, with $\gamma = 1.4$, gas constant $R = 1$, constant dynamic viscosity $\mu = 10^{-4}$ in the supplied native-64^3 protocol, Prandtl number $\mathrm{Pr} = 0.72$, and thermal conductivity $\kappa = \mu\gamma/[(\gamma - 1)\mathrm{Pr}]$. There is no external forcing. The finite-volume generator includes inviscid transport, Newtonian viscous stress with the Stokes hypothesis, viscous work, and Fourier heat conduction, so the integrals of mass, the three momentum components, and total energy are all conserved on the periodic domain. The corresponding invariant subspace is therefore full rank in the five represented channels, although it is still identified from training trajectories rather than prescribed from the equation name.

Initial density and pressure are positive lognormal transforms of smooth three-dimensional Gaussian random fields with independently randomized means; the velocity components use smooth random fields with random mean offsets and are rescaled so that the initial RMS Mach number is one. Consecutive stored frames are separated by $\Delta t = 0.02$. Reference trajectories are integrated in float64 using MUSCL reconstruction with an MC limiter, Rusanov inviscid face fluxes, centered viscous and thermal face fluxes, and SSPRK2 time integration. The internal step obeys acoustic and diffusive CFL bounds and is reduced on a rejected stage rather than repairing an inadmissible state by clipping. Stored trajectories are float32 arrays of shape $[N, T, 5, N_z, N_y, N_x]$.

C.2 Trajectory Sampling and Data Splits

For each of the three two-dimensional benchmarks, each seed generates 64 training, 16 validation, and 16 test trajectories. The three splits are drawn from independent random streams derived from the seed and a fixed per-split offset, so they never share initial conditions, and changing the seed regenerates the reference data as well as reinitializing the model. Every model within a seed sees identical trajectories. Generated caches are validated against a signature containing the complete configuration and physics-parameter record, so a reused cache cannot silently correspond to different physics. All two-dimensional models observe $p = 2$ context frames and are then rolled out autoregressively over the remaining frames of the trajectory, giving the rollout lengths in Table 6. Predictions are never reset to reference states during evaluation. The supplied native-64^3 three-dimensional continuation likewise uses $p = 2$, with six training, two validation, and two test trajectories of eight frames each; the three splits use the same independent deterministic random-stream construction.

For the two-dimensional benchmarks, the discovery and screening splits of Appendix A.1 are formed inside the training partition alone: of the 64 training trajectories, $\lfloor 64/5 \rfloor = 12$ are reserved for screening and the remaining 52 are used for discovery. This yields 2444 discovery and 564 screening rows of integrated changes for the fluid–phase-field benchmark and 1196 and 276 for shallow water. No validation or test trajectory enters the identification of C. Under this protocol the recovered subspaces agree with the reference physics: $(0, 0, 1)$ for the phase field, $(1, 0, 0)$ for shallow-water mass, and a three-dimensional porous subspace spanned by the two reaction-pair directions together with the static-porosity direction $(0, 0, 0, 1)$. The rank is discovered rather than prescribed. This agreement is a property of the evaluated ensembles and does not by itself establish population-level identifiability, as noted in Appendix A.1. In the native-64^3 continuation, five of the six training trajectories are used for discovery and one for screening; with eight frames per trajectory this gives 35 discovery and 7 screening integrated changes. For unforced compressible Navier–Stokes, recovery

of five independent invariant directions is physically expected, and the implementation treats this full-rank case as valid rather than automatically discarding it.

No input or output normalization is applied to the data. Every compared model predicts a residual increment added to the most recent observed frame, and normalization enters only through the loss, via the per-channel standard deviation of one-step training increments floored at 10^{-4} as defined in Appendix A.5. For the porous benchmark the divisor of the static porosity channel is instead fixed to 0.05, since its exactly zero increment variance would otherwise dominate the objective.

D Baselines and Experimental Protocol

D.1 Baseline Implementations

For the three two-dimensional benchmarks, all baselines are compact, self-contained reimplementations trained from scratch under the identical forecasting protocol; none uses pretrained weights or a multiphysics pretraining stage. Each predicts a residual increment from the same $p = 2$ context frames and is rolled out autoregressively. Table 7 lists the architectural settings, which are fixed a priori and never tuned; only optimization hyperparameters are searched (Appendix D.3).

Table 7: Baseline architecture settings for the three two-dimensional benchmarks, shared apart from the input and output channel counts implied by d.

Model	Configuration
U-Net (Ronneberger et al., 2015)	base width 8, two downsampling and two upsampling stages, GroupNorm and GELU
ConvLSTM (Shi et al., 2015)	hidden width 20, two stacked cells, circular 3×3 convolutions
FNO (Li et al., 2021)	width 20, 5 retained modes, depth 3, one-corner real FFT
PINO (Li et al., 2024)	the FNO architecture verbatim; differs only through the residual loss
LaMO (Tiwari et al., 2025)	embedding dimension 35, depth 2, patch size 4, zero-initialized head
CoDA-NO (Rahman et al., 2024)	width 11, 8 modes, depth 3, 2 encoding channels, 2 attention heads
clawNO-style (Liu et al., 2024)	width 13, 7 modes, depth 3; velocity scale 0.45, diffusion scale 0.16
NFV-style (Lichtlé et al., 2025)	width 105, depth 6; flux scale 0.50, step bound 0.60

The two transport controls deserve explicit description, since their structural constraints drive the interpretation in Section 4.2. The clawNO-style model, following Liu et al. (2024), predicts a stream function through a spectral encoder and uses its skew gradient as a bounded transport velocity, advancing every field by a conservative periodic flux update. The NFV-style model learns numerical face fluxes directly from local space–time stencils through a pointwise network and applies them in a periodic finite-volume update, with no weak-form or residual objective. In both cases the update is a difference of face fluxes for *every* channel independently, which is exactly the fieldwise conservation constraint discussed in Section 4.2; it is stronger than the selective balance structure of these benchmarks. LaMO and clawNO additionally append two coordinate channels to their inputs, whereas BalanceNO appends none (Appendix A.2).

Capacities are matched by trainable real scalar count, with one complex weight counted as two real scalars. Across the two-dimensional comparisons the budgets span 53.6k to 62.4k parameters, a spread of under $\pm 8\%$ about the mean; the per-benchmark counts appear in Table 10. Because architecture is excluded from the two-dimensional search, this matching holds for every reported cell and is not an artifact of a particular tuned configuration.

The three-dimensional code uses native 3D counterparts of the same model families rather than the two-dimensional modules in Table 7. A common target of $250{,}000$ trainable real scalars is supplied to the 3D model factory, which adjusts width parameters without adding inactive padding parameters; complex Fourier coefficients count as two real scalars. This parameter-count matching does not imply equal FLOPs or effective approximation capacity. BalanceNO's selected 3D settings are given in Appendix A.6.

D.2 Physics Residual of the PINO Baseline

PINO (Li et al., 2024) shares the FNO architecture of Li et al. (2021) and adds a collocation residual to the training objective. The time derivative is the second-order backward difference

$$\partial_t u \approx \frac{3\widehat{u}^{n+1} - 4u^n + u^{n-1}}{2\Delta t},$$

which is why the residual requires $p \geq 2$ context frames. The residual is expressed in state-increment units and divided by the same per-channel increment standard deviation as the supervised loss, so the two terms are commensurate, and it enters with weight 0.1. The right-hand sides follow Equations equation 10–equation 12 evaluated spectrally on the predicted state, with three implementation-specific details. For the fluid–phase-field system the momentum residual is Leray-projected and an additional incompressibility penalty $0.1\,\mathbb{E}[(\Delta t\,\nabla\cdot\mathbf{u}/u_{\mathrm{rms}})^2]$ is included. For shallow water the height is floored at 10^{-4} where it appears in a denominator, but not elsewhere. For the porous system the porosity is floored at 0.05 and the residual on the static porosity channel is set to zero.

For the three-dimensional compressible Navier–Stokes case, PINO uses the same BDF2 temporal construction but evaluates a continuous conservative compressible-Navier–Stokes right-hand side on the prediction using spectral spatial derivatives. The residual is formed directly in state-increment units,

$$\frac{3(\widehat{u}^{n+1} - u^n) - (u^n - u^{n-1})}{2} - \Delta t\,\mathcal{F}_{\mathrm{CNS}}(\widehat{u}^{n+1}),$$

and is normalized by the same channelwise increment scales as the supervised loss. A density floor of 10^{-6} is used only in denominators inside this loss. This residual is separate from the MUSCL–Rusanov data generator and is not used by BalanceNO.

D.3 Optimization Protocol and Configuration Selection

Within each two-dimensional benchmark, every compared model is trained under the same schedule, listed in Table 8. Optimization uses AdamW (Loshchilov & Hutter, 2019) with cosine annealing (Loshchilov & Hutter, 2017) over the full epoch budget, gradient-norm clipping at 1.0, deterministic cuDNN kernels, and a seeded shuffling generator. The rollout objective is the detached pushforward loss (Brandstetter et al., 2022) of Appendix A.5 with horizon H and per-step weights $1 + 0.25h$. Checkpoints are selected by validation rollout macro relative L^2 error. Positivity limiting is applied to the height field for shallow water and to the three species-storage fields for the porous benchmark, and to no field otherwise.

Table 8: Per-benchmark training protocol. The base learning rate d and rollout-loss horizon H also parameterize the search space in Table 9.

Benchmark	Epochs	Frames	Horizon H	Base rate d
Navier–Stokes–Cahn–Hilliard	24	48	5	3×10^{-3}
Rotating shallow water	20	24	5	2×10^{-3}
Heterogeneous porous reaction–diffusion	48	12	3	3×10^{-3}

Table 9 defines the twelve-point optimization search space. Identifiers 0 through 8 form the grid of three learning rates against three batch sizes at fixed weight decay and horizon; the remaining three probe a larger weight decay, a longer rollout-loss horizon, and their combination with the largest learning rate. Identifier 4 reproduces the protocol default. Epochs and architecture are not searched.

Table 9: The twelve optimization configurations, expressed through the per-benchmark base rate d and horizon H of Table 8. Identifier 4 is the protocol default.

Identifier	Learning rate	Batch size	Weight decay	Rollout horizon
$0, 1, 2$	$d/3$	$8, 16, 32$	10^{-5}	H
$3, 4, 5$	d	$8, 16, 32$	10^{-5}	H
$6, 7, 8$	$3d$	$8, 16, 32$	10^{-5}	H
9	d	16	10^{-4}	H
10	d	16	10^{-5}	$H + 2$
11	$3d$	32	10^{-4}	$H + 2$

Selection proceeds by two-stage successive halving (Jamieson & Talwalkar, 2016) on seed 0. The first stage trains all three benchmarks, nine models, and twelve configurations at a reduced budget of $\max(4, \lceil \text{epochs}/3 \rceil)$ epochs, which is 8, 7, and 16 epochs respectively. The six best configurations per benchmark and model, together with the protocol default, advance to a second stage at the full epoch budget. The surviving configuration with the lowest validation rollout macro relative L^2 error is then retrained across the five reported seeds. Test metrics are used only for the reported comparison and never for selection. Two cells depart from this rule. On the fluid–phase-field benchmark, CoDA-NO and the clawNO-style control revert to the default configuration under a documented pilot-test fallback rather than using their stage-two winners. Because selection operates on seed 0 alone while the reported results average five seeds, the configuration identifiers should be read as a fixed protocol rather than as a per-seed optimum.

Three-dimensional continuation protocol. The supplied native-64^3 continuation is kept separate from the two-dimensional twelve-point configuration search. Its frozen Stage-A protocol uses six training, two validation, and two test trajectories with eight frames, $p = 2$, 36 epochs, batch size 2, learning rate 10^{-3}, one-step rollout loss, stride-three temporal windows, and a common target of $250{,}000$ trainable real parameters. To reduce training memory, the supervised state loss is sampled on a uniformly selected 16^3 region, while the global BalanceNO features, Fourier filtering, spatial means, and the autoregressive history update are evaluated on the complete periodic 64^3 volume. Validation uses complete-volume rollouts.

D.4 Evaluation Metrics

Let $\widehat{u}^{n+h}$ and u^{n+h} for $h = 1, \dots, T$ denote the predicted and reference states over the rollout, indexed by channel c and generic grid cell $\boldsymbol{\xi}$, and recall $m(u) = \sum_{\boldsymbol{\xi}} u_{\boldsymbol{\xi}}$ from Equation equation 1. All quantities below are computed per test trajectory and then averaged over trajectories and seeds. The two relative-error measures differ only in the order of aggregation:

$$\text{macro} = \frac{1}{d} \sum_{c=1}^{d} \frac{\left(\sum_{h,\boldsymbol{\xi}} (\widehat{u}^{n+h}_{\boldsymbol{\xi} c} - u^{n+h}_{\boldsymbol{\xi} c})^2 \right)^{1/2}}{\left(\sum_{h,\boldsymbol{\xi}} (u^{n+h}_{\boldsymbol{\xi} c})^2 \right)^{1/2}}, \qquad \text{aggregate} = \frac{\left(\sum_{h,c,\boldsymbol{\xi}} (\widehat{u} - u)^2 \right)^{1/2}}{\left(\sum_{h,c,\boldsymbol{\xi}} u^2 \right)^{1/2}}. \tag{13}$$

The macro form weights every field equally and is the quantity used for checkpoint and configuration selection; the aggregate form is dominated by the channels of largest magnitude. Reporting both is necessary because they can order models differently when channel scales differ, as observed for the fluid–phase-field benchmark in Table 1. Rollout MSE is the plain mean squared error over the same index set. The field-total error compares domain-integrated fields against the reference,

$$\text{field-total} = \frac{1}{dT} \sum_{c,h} \frac{\left| m_c(\widehat{u}^{n+h}) - m_c(u^{n+h}) \right|}{\sum_{\boldsymbol{\xi}} \left| u^{n+h}_{\boldsymbol{\xi} c} \right|}. \tag{14}$$

Invariant drift instead measures departure from the model's own starting state. With $P_C^{\perp} = I_d - P_C = C^{\top}(CC^{\top})^{\dagger}C$ the projector onto the conserved row space, and u^n the last observed context frame,

$$\text{drift} = \frac{1}{T} \sum_{h=1}^{T} \frac{\left\| P_C^{\perp} \left[m(\widehat{u}^{n+h}) - m(u^n) \right] \right\|_2}{\left\| \sum_{\boldsymbol{\xi}} |u^n_{\boldsymbol{\xi}}| \right\|_2}. \tag{15}$$

Equations equation 14 and equation 15 are deliberately distinct, and the distinction is what separates correct selective conservation from overconservation: a model that freezes every field total scores at roundoff on Equation equation 15 while necessarily incurring a large error in Equation equation 14 whenever a nonconserved component total changes in the reference solution. For rotating shallow water, mass drift is the h component of Equation equation 15 in absolute relative form, the integrated-momentum error is $\|(m_{hu}, m_{hv})_{\mathrm{pred}} - (m_{hu}, m_{hv})_{\mathrm{ref}}\|_2$ normalized by the reference momentum norm, and the energy error compares

$$E(u) = \sum_{ij} \left[\tfrac{(hu)^2+(hv)^2}{2\max(h,10^{-6})} + \tfrac{1}{2} g h^2 \right]$$

against the reference trajectory rather than against its initial value, so it is a state-prediction diagnostic and not a conservation test. The nonpositive-height fraction counts predicted cells with $h \leq 0$ over all rollout steps. For the porous benchmark, the two pair drifts are Equation equation 15 evaluated on the individual combinations $q_A + q_C$ and $q_B + q_C$, each normalized by its own initial total, and the nonpositive-storage fraction counts predicted cells with $q_A \leq 0$, $q_B \leq 0$, or $q_C \leq 0$. Finally, the one-step MSE columns are computed on the validation split, since that split also drives checkpoint and configuration selection; they indicate local approximation quality and are not independent test results.

For three-dimensional compressible Navier–Stokes, the code additionally reports the relative drift of each of the five physical field totals using float64 spatial reductions and the initial integrated absolute magnitude of the corresponding field as the denominator. It also reports the minimum predicted density, the minimum pressure reconstructed from the conservative state, and the fraction of predicted cells that are nonfinite or have nonpositive density or pressure. These diagnostics report admissibility and conservation directly from the unmodified 3D BalanceNO rollout.

D.5 Hardware, Software, and Computational Cost

All reported two-dimensional results were produced on a single NVIDIA L40S GPU per run, with eight CPU cores and 32 GB of host memory per job, on a SLURM-managed cluster; the five seeds of each benchmark and model pair ran as independent array tasks. Every one of the twenty-seven reported two-dimensional benchmark–model cells used this device, so the wall-time and memory figures in Table 10 are comparable across models. Wall time measures the training loop including per-epoch validation, and peak memory is the maximum allocated device memory over training and evaluation; the reported device percentage is allocated memory relative to device capacity, not utilization. The two-dimensional runs use PyTorch (Paszke et al., 2019) with deterministic cuDNN kernels, disabled autotuning, and a seeded data-shuffling generator. Reference trajectories are regenerable from the seed and split identifier alone, and each cached dataset is keyed to the full physics and configuration record described in Appendix C.1, so a cache produced under different parameters cannot be reused silently.

Table 10: Training cost, mean $\pm$ std over five seeds. The 2D and 3D blocks use separate parameter budgets (6×10^4 and 2.5×10^5) and protocols, so values are not comparable across blocks. Parameter counts are a declared budget and are not ranked.

	2D: NS–Cahn–Hilliard			3D: compressible NS (64^3)		
Model	params	GPU (MiB)	wall (s)	params	GPU (MiB)	wall (s)
BalanceNO	61,821	439.9	392.8 ± 1.5	250,904	11,968	$8{,}804 \pm 46.4$
CoDA-NO	62,124	3,010	605.2 ± 10.2	267,216	2,302	$76{,}108 \pm 248.3$
clawNO-style	55,811	392.8	137.4 ± 4.1	260,782	1,832	$1{,}380 \pm 20.9$
NFV-style	59,854	9,192	150.7 ± 0.35	248,433	13,091	$6{,}801 \pm 24.1$
LaMO	58,768	**129.3**	$2{,}145 \pm 46.8$	251,384	5,964	$6{,}104 \pm 16.0$
FNO	62,043	182.3	166.7 ± 1.9	260,642	**1,437**	$1{,}260 \pm 12.4$
PINO	62,043	197.1	383.5 ± 4.4	260,642	2,104	$1{,}523 \pm 20.0$
U-Net	53,747	367.2	109.7 ± 1.5	229,253	1,594	$2{,}250 \pm 18.4$
ConvLSTM	62,003	986.0	$\mathbf{67.0 \pm 0.17}$	246,266	6,336	$8{,}644 \pm 21.4$

Table 10 reports the computational costs. Parameter counts are matched within each block, so BalanceNO's accuracy advantage is not associated with a larger parameter budget. On the 2D benchmark, its wall time is lower than that of CoDA-NO and LaMO but higher than that of the remaining models, and its peak memory is lower than that of CoDA-NO, NFV-style, and ConvLSTM but higher than that of the others. In 3D, the local face heads make BalanceNO comparatively expensive: its wall time is second only to CoDA-NO, and its peak memory is second only to NFV-style.

E COMPLETE RESULTS

Tables 11–17 report every recorded metric for each benchmark as mean ± std over five seeds; the best mean in each column is bolded, together with any entry within one standard deviation of it; the main-text tables in Section 4 are drawn from them. The validation one-step MSE column is a selection quantity rather than an independent test result. For the nonpositivity fractions in Tables 13 and 15, BalanceNO applies a fixed positivity limiter (Appendix A.4), so its zero values partly reflect that limiter rather than the balance update alone.

Table 11: Navier–Stokes–Cahn–Hilliard: forecasting metrics.

Model	rollout macro rel-L^2	aggregate rel-L^2	rollout MSE	field-total rel. err.	val. one-step MSE
BalanceNO	$\mathbf{8.8 \times 10^{-2} \pm 4.8 \times 10^{-2}}$	$\mathbf{1.2 \times 10^{-1} \pm 7.5 \times 10^{-2}}$	$\mathbf{2.3 \times 10^{-3} \pm 2.9 \times 10^{-3}}$	$\mathbf{3.2 \times 10^{-2} \pm 2.8 \times 10^{-2}}$	$\mathbf{4.9 \times 10^{-6} \pm 4.9 \times 10^{-6}}$
CoDA-NO	$5.7 \times 10^{-1} \pm 7.1 \times 10^{-2}$	$9.6 \times 10^{-1} \pm 7.1 \times 10^{-2}$	$1.0 \times 10^{-1} \pm 1.7 \times 10^{-2}$	$1.1 \times 10^{-1} \pm 4.0 \times 10^{-2}$	$2.3 \times 10^{-4} \pm 7.9 \times 10^{-5}$
clawNO-style	$1.1 \times 10^{0} \pm 2.8 \times 10^{-2}$	$8.5 \times 10^{-1} \pm 1.2 \times 10^{-2}$	$7.9 \times 10^{-2} \pm 2.8 \times 10^{-3}$	$\mathbf{5.7 \times 10^{-2} \pm 2.6 \times 10^{-3}}$	$2.5 \times 10^{-4} \pm 8.8 \times 10^{-6}$
NFV-style	$6.3 \times 10^{1} \pm 6.2 \times 10^{0}$	$2.0 \times 10^{1} \pm 2.3 \times 10^{0}$	$4.2 \times 10^{1} \pm 9.7 \times 10^{0}$	$\mathbf{5.7 \times 10^{-2} \pm 2.6 \times 10^{-3}}$	$2.2 \times 10^{-4} \pm 3.6 \times 10^{-5}$
LaMO	$2.3 \times 10^{-1} \pm 1.2 \times 10^{-2}$	$3.5 \times 10^{-1} \pm 2.4 \times 10^{-2}$	$1.4 \times 10^{-2} \pm 2.1 \times 10^{-3}$	$\mathbf{4.4 \times 10^{-2} \pm 8.3 \times 10^{-3}}$	$\mathbf{7.7 \times 10^{-6} \pm 1.0 \times 10^{-6}}$
FNO	$3.6 \times 10^{-1} \pm 1.5 \times 10^{-2}$	$6.1 \times 10^{-1} \pm 3.5 \times 10^{-2}$	$4.2 \times 10^{-2} \pm 5.7 \times 10^{-3}$	$6.3 \times 10^{-2} \pm 2.9 \times 10^{-3}$	$8.8 \times 10^{-5} \pm 1.1 \times 10^{-5}$
PINO	$3.7 \times 10^{-1} \pm 1.5 \times 10^{-2}$	$6.1 \times 10^{-1} \pm 3.1 \times 10^{-2}$	$4.2 \times 10^{-2} \pm 5.1 \times 10^{-3}$	$6.8 \times 10^{-2} \pm 1.0 \times 10^{-2}$	$7.8 \times 10^{-5} \pm 7.5 \times 10^{-6}$
U-Net	$3.8 \times 10^{-1} \pm 5.6 \times 10^{-2}$	$5.4 \times 10^{-1} \pm 6.0 \times 10^{-2}$	$3.3 \times 10^{-2} \pm 7.4 \times 10^{-3}$	$6.1 \times 10^{-2} \pm 1.4 \times 10^{-2}$	$4.5 \times 10^{-5} \pm 2.0 \times 10^{-5}$
ConvLSTM	$4.4 \times 10^{-1} \pm 2.4 \times 10^{-2}$	$6.3 \times 10^{-1} \pm 2.5 \times 10^{-2}$	$4.3 \times 10^{-2} \pm 3.5 \times 10^{-3}$	$1.7 \times 10^{-1} \pm 2.0 \times 10^{-2}$	$4.9 \times 10^{-5} \pm 5.3 \times 10^{-6}$

Table 12: Rotating shallow water: forecasting metrics.

Model	rollout macro rel-L^2	aggregate rel-L^2	rollout MSE	field-total rel. err.	val. one-step MSE
BalanceNO	$\mathbf{2.2 \times 10^{-2} \pm 1.5 \times 10^{-3}}$	$\mathbf{5.7 \times 10^{-3} \pm 5.4 \times 10^{-4}}$	$\mathbf{1.2 \times 10^{-5} \pm 2.7 \times 10^{-6}}$	$\mathbf{8.3 \times 10^{-3} \pm 9.1 \times 10^{-4}}$	$\mathbf{8.9 \times 10^{-7} \pm 2.2 \times 10^{-7}}$
CoDA-NO	$6.4 \times 10^{-1} \pm 7.9 \times 10^{-3}$	$1.6 \times 10^{-1} \pm 9.9 \times 10^{-3}$	$9.1 \times 10^{-3} \pm 1.4 \times 10^{-3}$	$8.2 \times 10^{-1} \pm 4.5 \times 10^{-2}$	$9.5 \times 10^{-5} \pm 6.0 \times 10^{-6}$
clawNO-style	$7.7 \times 10^{-1} \pm 1.5 \times 10^{-2}$	$2.0 \times 10^{-1} \pm 7.5 \times 10^{-3}$	$1.4 \times 10^{-2} \pm 1.5 \times 10^{-3}$	$8.1 \times 10^{-1} \pm 5.5 \times 10^{-2}$	$3.7 \times 10^{-4} \pm 9.5 \times 10^{-6}$
NFV-style	$8.0 \times 10^{-1} \pm 1.5 \times 10^{-2}$	$2.0 \times 10^{-1} \pm 1.1 \times 10^{-2}$	$1.5 \times 10^{-2} \pm 1.9 \times 10^{-3}$	$8.1 \times 10^{-1} \pm 5.5 \times 10^{-2}$	$3.5 \times 10^{-4} \pm 1.5 \times 10^{-5}$
LaMO	$1.1 \times 10^{-1} \pm 1.4 \times 10^{-2}$	$3.1 \times 10^{-2} \pm 4.4 \times 10^{-3}$	$3.4 \times 10^{-4} \pm 1.0 \times 10^{-4}$	$3.0 \times 10^{-2} \pm 8.6 \times 10^{-3}$	$7.9 \times 10^{-6} \pm 9.0 \times 10^{-7}$
FNO	$1.2 \times 10^{-1} \pm 8.0 \times 10^{-3}$	$3.4 \times 10^{-2} \pm 7.3 \times 10^{-4}$	$4.0 \times 10^{-4} \pm 1.4 \times 10^{-5}$	$2.8 \times 10^{-2} \pm 2.7 \times 10^{-3}$	$3.4 \times 10^{-5} \pm 1.0 \times 10^{-5}$
PINO	$1.3 \times 10^{-1} \pm 7.1 \times 10^{-3}$	$3.5 \times 10^{-2} \pm 1.0 \times 10^{-3}$	$4.3 \times 10^{-4} \pm 3.6 \times 10^{-5}$	$2.8 \times 10^{-2} \pm 2.9 \times 10^{-3}$	$5.0 \times 10^{-5} \pm 1.0 \times 10^{-5}$
U-Net	$2.2 \times 10^{-1} \pm 1.3 \times 10^{-2}$	$5.8 \times 10^{-2} \pm 5.0 \times 10^{-3}$	$1.2 \times 10^{-3} \pm 2.3 \times 10^{-4}$	$1.1 \times 10^{-1} \pm 2.5 \times 10^{-2}$	$3.6 \times 10^{-5} \pm 8.2 \times 10^{-6}$
ConvLSTM	$2.9 \times 10^{-1} \pm 1.4 \times 10^{-2}$	$8.0 \times 10^{-2} \pm 4.4 \times 10^{-3}$	$2.2 \times 10^{-3} \pm 2.7 \times 10^{-4}$	$1.3 \times 10^{-1} \pm 8.5 \times 10^{-3}$	$2.4 \times 10^{-5} \pm 2.1 \times 10^{-6}$

Table 13: Rotating shallow water: balance and admissibility diagnostics.

Model	invariant drift	mass drift	integrated momentum rel-L^2	energy rel. error	nonpositive heights
BalanceNO	$\mathbf{2.9 \times 10^{-8} \pm 7.0 \times 10^{-9}}$	$\mathbf{3.0 \times 10^{-8} \pm 7.1 \times 10^{-9}}$	$\mathbf{9.2 \times 10^{-3} \pm 1.4 \times 10^{-3}}$	$\mathbf{2.8 \times 10^{-4} \pm 1.1 \times 10^{-4}}$	0
CoDA-NO	$9.1 \times 10^{-3} \pm 6.6 \times 10^{-3}$	$9.3 \times 10^{-3} \pm 6.7 \times 10^{-3}$	$8.7 \times 10^{-1} \pm 2.1 \times 10^{-2}$	$1.8 \times 10^{-2} \pm 1.3 \times 10^{-2}$	0
clawNO-style	$\mathbf{2.9 \times 10^{-8} \pm 8.1 \times 10^{-9}}$	$\mathbf{3.0 \times 10^{-8} \pm 8.2 \times 10^{-9}}$	8.6×10^{-1}	$2.2 \times 10^{-2} \pm 2.8 \times 10^{-2}$	$2.1 \times 10^{-6} \pm 4.2 \times 10^{-6}$
NFV-style	$\mathbf{2.8 \times 10^{-8} \pm 5.1 \times 10^{-9}}$	$\mathbf{2.9 \times 10^{-8} \pm 5.2 \times 10^{-9}}$	8.6×10^{-1}	$1.1 \times 10^{-2} \pm 1.2 \times 10^{-3}$	0
LaMO	$5.2 \times 10^{-3} \pm 1.0 \times 10^{-3}$	$5.3 \times 10^{-3} \pm 1.1 \times 10^{-3}$	$3.3 \times 10^{-2} \pm 9.3 \times 10^{-3}$	$1.1 \times 10^{-2} \pm 2.3 \times 10^{-3}$	0
FNO	$2.8 \times 10^{-3} \pm 5.8 \times 10^{-4}$	$2.8 \times 10^{-3} \pm 5.9 \times 10^{-4}$	$3.1 \times 10^{-2} \pm 3.5 \times 10^{-3}$	$5.9 \times 10^{-3} \pm 1.2 \times 10^{-3}$	0
PINO	$3.6 \times 10^{-3} \pm 9.6 \times 10^{-4}$	$3.7 \times 10^{-3} \pm 9.9 \times 10^{-4}$	$3.1 \times 10^{-2} \pm 4.7 \times 10^{-3}$	$7.3 \times 10^{-3} \pm 1.9 \times 10^{-3}$	0
U-Net	$5.9 \times 10^{-3} \pm 1.1 \times 10^{-3}$	$5.9 \times 10^{-3} \pm 1.2 \times 10^{-3}$	$1.2 \times 10^{-1} \pm 1.5 \times 10^{-2}$	$1.3 \times 10^{-2} \pm 2.8 \times 10^{-3}$	0
ConvLSTM	$2.3 \times 10^{-2} \pm 4.2 \times 10^{-3}$	$2.3 \times 10^{-2} \pm 4.2 \times 10^{-3}$	$1.5 \times 10^{-1} \pm 6.7 \times 10^{-3}$	$4.7 \times 10^{-2} \pm 8.5 \times 10^{-3}$	0

Table 14: Heterogeneous porous reaction–diffusion: forecasting metrics.

Model	rollout macro rel-L^2	aggregate rel-L^2	rollout MSE	field-total rel. err.	val. one-step MSE
BalanceNO	$\mathbf{1.1 \times 10^{-2} \pm 1.2 \times 10^{-3}}$	$\mathbf{7.2 \times 10^{-3} \pm 8.5 \times 10^{-4}}$	$\mathbf{8.2 \times 10^{-6} \pm 1.7 \times 10^{-6}}$	$\mathbf{1.0 \times 10^{-3} \pm 4.5 \times 10^{-4}}$	$\mathbf{5.2 \times 10^{-7} \pm 7.6 \times 10^{-8}}$
CoDA-NO	$9.3 \times 10^{-2} \pm 1.4 \times 10^{-2}$	$6.3 \times 10^{-2} \pm 9.4 \times 10^{-3}$	$6.4 \times 10^{-4} \pm 2.0 \times 10^{-4}$	$6.3 \times 10^{-2} \pm 9.8 \times 10^{-3}$	$3.2 \times 10^{-5} \pm 8.9 \times 10^{-6}$
clawNO-style	$2.9 \times 10^{-1} \pm 1.2 \times 10^{-2}$	$1.7 \times 10^{-1} \pm 7.0 \times 10^{-3}$	$4.8 \times 10^{-3} \pm 4.1 \times 10^{-4}$	$2.5 \times 10^{-1} \pm 1.3 \times 10^{-2}$	$1.4 \times 10^{-4} \pm 7.8 \times 10^{-6}$
NFV-style	$2.6 \times 10^{-1} \pm 1.1 \times 10^{-2}$	$1.6 \times 10^{-1} \pm 6.4 \times 10^{-3}$	$3.9 \times 10^{-3} \pm 3.5 \times 10^{-4}$	$2.5 \times 10^{-1} \pm 1.3 \times 10^{-2}$	$1.2 \times 10^{-4} \pm 4.5 \times 10^{-6}$
LaMO	$2.2 \times 10^{-2} \pm 8.5 \times 10^{-4}$	$1.4 \times 10^{-2} \pm 5.6 \times 10^{-4}$	$3.2 \times 10^{-5} \pm 3.3 \times 10^{-6}$	$2.6 \times 10^{-3} \pm 5.1 \times 10^{-4}$	$1.5 \times 10^{-6} \pm 1.3 \times 10^{-7}$
FNO	$3.3 \times 10^{-2} \pm 3.2 \times 10^{-3}$	$2.1 \times 10^{-2} \pm 2.2 \times 10^{-3}$	$6.8 \times 10^{-5} \pm 1.3 \times 10^{-5}$	$6.3 \times 10^{-3} \pm 9.2 \times 10^{-4}$	$3.4 \times 10^{-6} \pm 3.4 \times 10^{-7}$
PINO	$1.6 \times 10^{-1} \pm 1.2 \times 10^{-2}$	$1.4 \times 10^{-1} \pm 8.8 \times 10^{-3}$	$3.0 \times 10^{-3} \pm 4.0 \times 10^{-4}$	$6.9 \times 10^{-2} \pm 1.8 \times 10^{-2}$	$9.9 \times 10^{-4} \pm 1.1 \times 10^{-4}$
U-Net	$4.0 \times 10^{-2} \pm 5.1 \times 10^{-3}$	$2.5 \times 10^{-2} \pm 3.4 \times 10^{-3}$	$1.0 \times 10^{-4} \pm 2.8 \times 10^{-5}$	$1.5 \times 10^{-2} \pm 5.9 \times 10^{-3}$	$5.7 \times 10^{-6} \pm 6.9 \times 10^{-7}$
ConvLSTM	$2.2 \times 10^{-2} \pm 1.0 \times 10^{-3}$	$1.4 \times 10^{-2} \pm 8.4 \times 10^{-4}$	$3.1 \times 10^{-5} \pm 3.1 \times 10^{-6}$	$5.6 \times 10^{-3} \pm 6.5 \times 10^{-4}$	$1.9 \times 10^{-6} \pm 2.3 \times 10^{-7}$

Table 15: Heterogeneous porous reaction–diffusion: balance and admissibility diagnostics.

Model	invariant drift	pair q_A+q_C drift	pair q_B+q_C drift	nonpositive storage
BalanceNO	$\mathbf{4.6 \times 10^{-8} \pm 8.0 \times 10^{-9}}$	$3.4 \times 10^{-8} \pm 3.5 \times 10^{-9}$	$3.0 \times 10^{-8} \pm 3.9 \times 10^{-9}$	0
CoDA-NO	$3.0 \times 10^{-2} \pm 4.0 \times 10^{-3}$	$4.1 \times 10^{-2} \pm 6.9 \times 10^{-3}$	$4.6 \times 10^{-2} \pm 1.3 \times 10^{-2}$	$4.9 \times 10^{-4} \pm 9.5 \times 10^{-4}$
clawNO-style	$\mathbf{4.4 \times 10^{-8} \pm 5.8 \times 10^{-9}}$	$\mathbf{2.6 \times 10^{-8} \pm 2.7 \times 10^{-9}}$	$\mathbf{2.7 \times 10^{-8} \pm 2.6 \times 10^{-9}}$	0
NFV-style	$\mathbf{4.4 \times 10^{-8} \pm 8.5 \times 10^{-9}}$	$\mathbf{2.7 \times 10^{-8} \pm 4.9 \times 10^{-9}}$	$\mathbf{2.9 \times 10^{-8} \pm 7.4 \times 10^{-9}}$	$4.9 \times 10^{-3} \pm 1.4 \times 10^{-3}$
LaMO	$1.6 \times 10^{-3} \pm 3.7 \times 10^{-4}$	$2.4 \times 10^{-3} \pm 3.8 \times 10^{-4}$	$2.4 \times 10^{-3} \pm 7.7 \times 10^{-4}$	0
FNO	$2.9 \times 10^{-3} \pm 5.8 \times 10^{-4}$	$3.9 \times 10^{-3} \pm 7.5 \times 10^{-4}$	$4.2 \times 10^{-3} \pm 1.5 \times 10^{-3}$	0
PINO	$6.7 \times 10^{-2} \pm 1.1 \times 10^{-2}$	$5.0 \times 10^{-2} \pm 1.6 \times 10^{-2}$	$4.6 \times 10^{-2} \pm 1.8 \times 10^{-2}$	0
U-Net	$2.8 \times 10^{-3} \pm 1.5 \times 10^{-3}$	$4.0 \times 10^{-3} \pm 1.7 \times 10^{-3}$	$4.1 \times 10^{-3} \pm 3.2 \times 10^{-3}$	$5.8 \times 10^{-6} \pm 7.2 \times 10^{-6}$
ConvLSTM	$3.6 \times 10^{-3} \pm 5.1 \times 10^{-4}$	$6.4 \times 10^{-3} \pm 2.8 \times 10^{-4}$	$5.0 \times 10^{-3} \pm 1.3 \times 10^{-3}$	0

Table 16: 3D compressible Navier–Stokes (64^3): forecasting metrics.

Model	rollout macro rel-L^2	aggregate rel-L^2	rollout MSE	field-total rel. err.	final-step rel-L^2	val. one-step MSE
BalanceNO	$\mathbf{1.0 \times 10^{-1} \pm 9.6 \times 10^{-3}}$	$\mathbf{7.0 \times 10^{-2} \pm 5.8 \times 10^{-3}}$	$\mathbf{9.6 \times 10^{-3} \pm 1.5 \times 10^{-3}}$	$1.5 \times 10^{-8} \pm 1.5 \times 10^{-9}$	$\mathbf{1.3 \times 10^{-1} \pm 1.9 \times 10^{-2}}$	$\mathbf{7.4 \times 10^{-4} \pm 7.1 \times 10^{-5}}$
CoDA-NO	$4.7 \times 10^{-1} \pm 1.3 \times 10^{-2}$	$3.7 \times 10^{-1} \pm 5.4 \times 10^{-3}$	$2.5 \times 10^{-1} \pm 1.1 \times 10^{-2}$	$1.1 \times 10^{-2} \pm 4.6 \times 10^{-3}$	$5.0 \times 10^{-1} \pm 5.7 \times 10^{-3}$	$2.2 \times 10^{-2} \pm 3.7 \times 10^{-3}$
clawNO-style	$6.5 \times 10^{-1} \pm 1.1 \times 10^{-1}$	$5.5 \times 10^{-1} \pm 9.8 \times 10^{-2}$	$5.8 \times 10^{-1} \pm 2.0 \times 10^{-1}$	$\mathbf{1.3 \times 10^{-8} \pm 1.0 \times 10^{-9}}$	$1.4 \times 10^{0} \pm 3.2 \times 10^{-1}$	$1.9 \times 10^{-2} \pm 2.5 \times 10^{-3}$
NFV-style	$1.5 \times 10^{0} \pm 2.0 \times 10^{-1}$	$1.9 \times 10^{0} \pm 3.9 \times 10^{-1}$	$6.7 \times 10^{0} \pm 2.8 \times 10^{0}$	$\mathbf{1.4 \times 10^{-8} \pm 1.5 \times 10^{-9}}$	$5.2 \times 10^{0} \pm 1.3 \times 10^{0}$	$1.5 \times 10^{-2} \pm 1.8 \times 10^{-3}$
LaMO	$2.2 \times 10^{-1} \pm 4.1 \times 10^{-3}$	$1.8 \times 10^{-1} \pm 3.2 \times 10^{-3}$	$6.1 \times 10^{-2} \pm 2.3 \times 10^{-3}$	$1.7 \times 10^{-2} \pm 9.7 \times 10^{-4}$	$3.0 \times 10^{-1} \pm 4.5 \times 10^{-3}$	$2.8 \times 10^{-3} \pm 2.9 \times 10^{-4}$
FNO	$3.2 \times 10^{-1} \pm 7.6 \times 10^{-3}$	$2.4 \times 10^{-1} \pm 6.0 \times 10^{-3}$	$1.1 \times 10^{-1} \pm 6.4 \times 10^{-3}$	$2.8 \times 10^{-2} \pm 2.6 \times 10^{-3}$	$3.4 \times 10^{-1} \pm 1.0 \times 10^{-2}$	$1.9 \times 10^{-2} \pm 1.6 \times 10^{-3}$
PINO	$3.1 \times 10^{-1} \pm 1.2 \times 10^{-2}$	$2.4 \times 10^{-1} \pm 9.1 \times 10^{-3}$	$1.1 \times 10^{-1} \pm 9.6 \times 10^{-3}$	$2.5 \times 10^{-2} \pm 1.9 \times 10^{-3}$	$3.4 \times 10^{-1} \pm 8.4 \times 10^{-3}$	$1.9 \times 10^{-2} \pm 1.8 \times 10^{-3}$
U-Net	$1.8 \times 10^{-1} \pm 4.1 \times 10^{-3}$	$1.4 \times 10^{-1} \pm 4.3 \times 10^{-3}$	$3.8 \times 10^{-2} \pm 2.2 \times 10^{-3}$	$2.1 \times 10^{-2} \pm 1.5 \times 10^{-3}$	$2.1 \times 10^{-1} \pm 6.3 \times 10^{-3}$	$5.5 \times 10^{-3} \pm 6.5 \times 10^{-4}$
ConvLSTM	$1.3 \times 10^{-1} \pm 4.7 \times 10^{-3}$	$1.0 \times 10^{-1} \pm 3.1 \times 10^{-3}$	$2.0 \times 10^{-2} \pm 1.2 \times 10^{-3}$	$6.3 \times 10^{-3} \pm 8.6 \times 10^{-4}$	$1.5 \times 10^{-1} \pm 3.8 \times 10^{-3}$	$2.9 \times 10^{-3} \pm 3.8 \times 10^{-4}$

Table 17: 3D compressible Navier–Stokes (64^3): drift of the invariant subspace and of each conservative field total $[\rho,\ \rho u,\ \rho v,\ \rho w,\ E]$.

Model	invariant drift	mass	x-momentum	y-momentum	z-momentum	energy
BalanceNO	$\mathbf{3.4 \times 10^{-8} \pm 8.9 \times 10^{-9}}$	$2.5 \times 10^{-10} \pm 2.0 \times 10^{-10}$	$5.4 \times 10^{-10} \pm 4.6 \times 10^{-10}$	$5.8 \times 10^{-10} \pm 2.1 \times 10^{-10}$	$5.2 \times 10^{-10} \pm 2.8 \times 10^{-10}$	$8.9 \times 10^{-10} \pm 5.2 \times 10^{-10}$
CoDA-NO	$1.4 \times 10^{-2} \pm 5.8 \times 10^{-3}$	$7.4 \times 10^{-3} \pm 4.8 \times 10^{-3}$	$9.2 \times 10^{-3} \pm 4.8 \times 10^{-3}$	$9.1 \times 10^{-3} \pm 4.1 \times 10^{-3}$	$8.6 \times 10^{-3} \pm 4.1 \times 10^{-3}$	$1.5 \times 10^{-2} \pm 6.6 \times 10^{-3}$
clawNO-style	$\mathbf{3.4 \times 10^{-8} \pm 8.2 \times 10^{-9}}$	$2.7 \times 10^{-10} \pm 3.3 \times 10^{-11}$	$3.0 \times 10^{-10} \pm 2.1 \times 10^{-11}$	$3.4 \times 10^{-10} \pm 6.0 \times 10^{-11}$	$7.0 \times 10^{-10} \pm 1.1 \times 10^{-10}$	$4.0 \times 10^{-10} \pm 1.1 \times 10^{-10}$
NFV-style	$\mathbf{3.4 \times 10^{-8} \pm 7.6 \times 10^{-9}}$	$\mathbf{1.5 \times 10^{-10} \pm 5.2 \times 10^{-12}}$	$\mathbf{1.8 \times 10^{-10} \pm 5.6 \times 10^{-11}}$	$\mathbf{1.7 \times 10^{-10} \pm 3.7 \times 10^{-11}}$	$\mathbf{1.6 \times 10^{-10} \pm 4.0 \times 10^{-11}}$	$\mathbf{1.6 \times 10^{-10} \pm 1.9 \times 10^{-11}}$
LaMO	$1.4 \times 10^{-2} \pm 2.3 \times 10^{-3}$	$9.2 \times 10^{-3} \pm 3.2 \times 10^{-3}$	$1.8 \times 10^{-2} \pm 3.8 \times 10^{-3}$	$1.7 \times 10^{-2} \pm 4.2 \times 10^{-3}$	$2.1 \times 10^{-2} \pm 3.5 \times 10^{-3}$	$1.2 \times 10^{-2} \pm 3.4 \times 10^{-3}$
FNO	$2.3 \times 10^{-2} \pm 7.6 \times 10^{-3}$	$1.4 \times 10^{-2} \pm 6.8 \times 10^{-3}$	$3.1 \times 10^{-2} \pm 8.2 \times 10^{-3}$	$3.0 \times 10^{-2} \pm 3.3 \times 10^{-3}$	$3.1 \times 10^{-2} \pm 4.2 \times 10^{-3}$	$2.1 \times 10^{-2} \pm 9.5 \times 10^{-3}$
PINO	$2.1 \times 10^{-2} \pm 2.8 \times 10^{-3}$	$1.2 \times 10^{-2} \pm 1.7 \times 10^{-3}$	$2.7 \times 10^{-2} \pm 1.9 \times 10^{-3}$	$2.8 \times 10^{-2} \pm 5.2 \times 10^{-3}$	$2.7 \times 10^{-2} \pm 4.0 \times 10^{-3}$	$1.9 \times 10^{-2} \pm 3.7 \times 10^{-3}$
U-Net	$1.5 \times 10^{-2} \pm 3.2 \times 10^{-3}$	$6.7 \times 10^{-3} \pm 3.1 \times 10^{-3}$	$2.5 \times 10^{-2} \pm 9.7 \times 10^{-3}$	$2.7 \times 10^{-2} \pm 4.4 \times 10^{-3}$	$2.4 \times 10^{-2} \pm 6.5 \times 10^{-3}$	$1.1 \times 10^{-2} \pm 4.1 \times 10^{-3}$
ConvLSTM	$7.1 \times 10^{-3} \pm 8.3 \times 10^{-4}$	$2.3 \times 10^{-3} \pm 3.7 \times 10^{-4}$	$5.9 \times 10^{-3} \pm 8.5 \times 10^{-4}$	$6.3 \times 10^{-3} \pm 1.2 \times 10^{-3}$	$6.8 \times 10^{-3} \pm 2.5 \times 10^{-3}$	$7.3 \times 10^{-3} \pm 8.9 \times 10^{-4}$

F CONTROLLED ABLATIONS AND DECODER-COORDINATE COMPLETION

We isolate the contributions of selective conservation, conservative transport, and auxiliary transfer preconditioning while holding the learned feature extractor fixed. The experiment is deliberately a mechanistic diagnostic on a new synthetic system rather than a comparison with external state-of-the-art methods. All variants are trained from scratch and evaluated using the same trajectory splits and forecast protocol within each study. We retain an initial unsuccessful test round and report a subsequent, explicitly declared decoder revision on fresh confirmation seeds; the two cohorts are not pooled.

F.1 WEIGHTED EXCHANGE–ADVECTION–DIFFUSION BENCHMARK

Consider three real-valued fields $u = (a, b, z)^\top$ on the periodic domain $[0, 2\pi)^2$, governed by

$$\partial_t u_c + v_c \cdot \nabla u_c = \kappa_c \Delta u_c + (Au)_c, \qquad A = \begin{pmatrix} -0.60 & 0.70 & 0 \\ 0.30 & -0.35 & 0 \\ 0.10 & -0.20 & -0.20 \end{pmatrix}. \tag{16}$$

The velocities are $v_a = (0.8, 0.3)$, $v_b = (-0.35, 0.6)$ and $v_z = (0.25, -0.55)$, and the diffusivities are $(\kappa_a, \kappa_b, \kappa_z) = (0.04, 0.025, 0.06)$. Because $C_\star A = 0$ for $C_\star = (1, 2, 0)$ and A has rank two, the integrated dynamics have exactly one independent linear invariant,

$$\mathcal{I}_\star(u) = \int_{[0,2\pi)^2} (a + 2b)\, \mathrm{d}x. \tag{17}$$

Individual field totals nevertheless evolve. Furthermore, the conserved density $a+2b$ changes locally through transport. Consequently, conserving every field total or fixing the conserved density pointwise imposes an incorrect additional restriction. The third field supplies a nonconserved relaxation mode, and the nonunit coefficient in $C_\star$ tests discovery in physical field coordinates.

Reference trajectories are computed independently of BalanceNO on a 16×16 grid with observation interval $\Delta t = 0.15$. For each Fourier wavevector k, the update is

$$\widehat{u}^{n+1}(k) = \exp\big\{\Delta t \left[A - \mathrm{diag}_c\big(\kappa_c |k|^2 + \mathrm{i}\, v_c \cdot k\big)\right]\big\}\, \widehat{u}^n(k). \tag{18}$$

Thus the represented band-limited linear dynamics are advanced by their exact semigroup, up to floating-point evaluation of the matrix exponential and FFT; there is no finite-step integration truncation error or use of the BNO decoder in label generation. Reference arithmetic is float64 before conversion to float32 observations. Each initial field is a spatial mean drawn uniformly from $[0.2, 1.2]$ plus an independent Gaussian random field filtered by $\exp(-0.15|k|^2)\mathbf{1}_{0<|k|^2\le 9}$, normalized to unit spatial standard deviation and multiplied by an amplitude drawn uniformly from $[0.15, 0.35]$. Fields are real-valued; positivity is not imposed.

For each of five fresh confirmation seeds (200–204), we generate 24 training, eight validation and 16 test trajectories with independent split-specific random streams. Training trajectories contain 14 frames; validation and test trajectories contain 42. Two observed frames initialize each rollout, followed by either ten or 40 forecast steps. The 40-step horizon extends beyond the duration of the training trajectories, although full-horizon validation is used for model selection. Discovery uses 18 training trajectories and screening uses the other six, with thresholds 10^{-5} and 2×10^{-5}, respectively. All 24 are subsequently available for supervised neural training. Neither model-selection validation nor test trajectories enter invariant discovery. The known $C_\star$ is used only for physical diagnostics, not supplied to the model.

F.2 CAPACITY-CONTROLLED INTERVENTIONS

Let P be the projector onto the kernel of the discovered matrix, let $Q = I - P$, and let $\langle\cdot\rangle$ denote the spatial mean. The common network predicts coordinate fields (J_x, J_y, r, a). Write

$$\mathcal{T}(J, a) = -\mathcal{D}\big(J + \mathcal{R}\mathcal{B}a\big), \qquad -\mathcal{D}\mathcal{R}b = b - \langle b\rangle,$$

where $\mathcal{D}$ is the original periodic backward-difference divergence, $\mathcal{B}$ is the original two-sided low-pass filter and $\mathcal{R}$ is its cumulative-sum right-inverse construction. The six compared increments

$\delta u = u^{n+1} - u^n$ are

$$\begin{aligned}
\text{Full BNO:} &\quad \mathcal{T}(J, a + Qr) + Pr,\\
\text{No projection:} &\quad \mathcal{T}(J, a) + r,\\
\text{Global projection:} &\quad \mathcal{T}(J, a) + r - Q\langle r\rangle,\\
\text{Divergence only:} &\quad \mathcal{T}(J, a + r),\\
\text{Projection only:} &\quad \tfrac{1}{2}P(J_x + J_y + r + a),\\
\text{No preconditioning:} &\quad -\mathcal{D}\big((J_x + b)/\sqrt{2}, (J_y + b)/\sqrt{2}\big) + Pr, \quad b = a + Qr.
\end{aligned} \tag{19}$$

The completed full decoder routes the complementary interaction Qr into conservative transfers rather than discarding it pointwise. This coordinate completion is already present in the supplied 3D decoder; here it is adapted to two spatial directions. It changes neither the learned head budget nor the defining transport–interaction balance law. At full invariant rank, $P = 0$ and the full and divergence-only routes coincide. The projection-only routing retains the supplied four-group parameter-matched combination. The global-projection control preserves the same integrated invariant. Writing $\mathcal{Z}b = b - \langle b\rangle$, the completed full update differs from this control by exactly $\mathcal{Z}(\mathcal{B} - I)Qr$. They are therefore algebraically identical when $\mathcal{B} = I$. With the present low-pass filter, their comparison tests band-limited conservative routing of the complementary interaction, not a distinct global conservation law. The final intervention removes the auxiliary low-pass/right-inverse path and reuses its head as raw face coordinates; it does not isolate filtering from inverse-divergence conditioning.

Every variant uses the same two-layer circular ConvLSTM branch of width six, a width-four spectral branch with one residual Fourier block and four retained modes, and the same fusion and output heads, totaling 6,816 real trainable scalars (a complex scalar counts as two). All learned heads remain connected, although conservative flux gauges and projected directions naturally have unidentifiable components. There are no coordinate channels, positivity repairs, artificial diffusion, conservation penalties or PDE-residual losses. With coordinate completion disabled, the compact implementation reproduces the original 2D full, divergence-only and projection-only outputs and all parameter gradients with zero observed discrepancy at matched weights in the executed parity test. The discovery projector also matches. Additional tests verify the completed decoder against its algebraic formula, exact full-rank full/divergence equivalence, conservation, and gradient connectivity. Thus the completed configuration is explicitly distinguished from an unchanged 2D port.

An additional trained oracle-invariant model is unnecessary for this exact-data check: all five training splits recover one invariant, and the maximum float64 Frobenius discrepancy between the recovered row-space projector and that of $C_\star$ is 2.10×10^{-9}. The smallest normalized singular RMS among the two rejected nonconserved directions is 2.75×10^{-2}, whereas the largest accepted invariant singular RMS is 2.65×10^{-9}. These diagnostics verify recovery for this ensemble, not robustness to noisy or insufficiently excited observations. Backbone ablations are omitted because the backbone is deliberately fixed in the causal comparison.

F.3 Optimization and reporting protocol

The loss is the supplied detached two-step pushforward objective,

$$\mathcal{L} = \frac{1}{2}\sum_{h=0}^{1}(1 + 0.25h)\frac{1}{3G}\sum_{c,\xi}\left(\frac{\widehat{u}_{c,\xi}^{n+h+1} - u_{c,\xi}^{n+h+1}}{\max(\mathrm{std}_{\mathrm{train}}(\Delta u_c), 10^{-4})}\right)^2. \tag{20}$$

The prediction is detached before being appended to the next context window. We use AdamW, weight decay 10^{-5}, gradient-norm clipping at one, batch size 16 and training-window stride three, yielding 96 windows and six minibatches per epoch. Learning rates are selected separately for each variant from $\{3 \times 10^{-3}, 6 \times 10^{-3}, 10^{-2}\}$ using 60-epoch runs on two development seeds (100 and 101), with selection by their mean validation score. The selected rates are BalanceNO (full): 1.00×10^{-2}; No projection: 1.00×10^{-2}; Global projection: 6.00×10^{-3}; Divergence only: 1.00×10^{-2}; Projection only: 6.00×10^{-3}; No preconditioning: 1.00×10^{-2}. Each selected configuration is then trained for 100 epochs on five new seeds, changing both the data and initialization. Cosine learning-rate decay ends at five percent of the initial rate. Validation macro relative L^2 error over 40 steps selects a checkpoint every five epochs.

Initial result and declared protocol amendment. The first study used the original 2D routing $\mathcal{T}(J, a) + Pr$. A short validation pilot was uniformly extended from 30 to 60 epochs and from two to three learning rates, after which five models per variant were trained for 100 epochs on seeds 0–4. Although full BNO led the extended validation pilot, it did not lead the held-out primary endpoint: its mean 40-step macro error was 4.06%, compared with 3.79% for global projection. Table 19 retains that unsuccessful result.

After observing this failure, we declared one additional study using the complementary-interaction routing above, already motivated by the supplied 3D implementation. Hyperparameters were re-selected on development validation trajectories, and confirmation used new data and initialization seeds 200–204. The PDE, initial-condition distribution, grid, split sizes, training loss, learned architecture and forecast horizons were unchanged. All 30 revised checkpoints were frozen and hashed before access to their new test trajectories. This is an adaptive second study, not a claim that the original study was successful or that the full development process was preregistered. The original and revised cohorts must not be compared as a paired estimate of the coordinate-completion effect. Both studies and all validation failures remain in the reproduction archive.

For horizon H, the primary accuracy measure is the channel-macro relative L^2 error, computed per trajectory as

$$E_{\mathrm{macro}}(H) = \frac{1}{3} \sum_{c} \frac{\|\widehat{u}_c - u_c\|_{2,\{1,\dots,H\}\times G}}{\|u_c\|_{2,\{1,\dots,H\}\times G}}. \tag{21}$$

Aggregate error takes one relative norm over time, space and channels. Writing $m(u) = \sum_{\xi} u_{\xi}$ and $Q_\star = C_\star^\top (C_\star C_\star^\top)^{-1} C_\star$, the field-total error and invariant drift are

$$E_{\mathrm{total}} = \frac{1}{3H} \sum_{h,c} \frac{|m_c(\widehat{u}^h) - m_c(u^h)|}{\sum_{\xi} |u^h_{c,\xi}|}, \tag{22}$$

$$E_{\mathrm{inv}} = \frac{1}{H} \sum_{h} \frac{\|Q_\star[m(\widehat{u}^h) - m(u^0)]\|_2}{\|\sum_{\xi} |u^0_{\xi}|\|_2}. \tag{23}$$

Here u^0 is the last observed context frame. Reductions are evaluated in float64. We first average over the 16 test trajectories within a seed and then report the mean and sample standard deviation of the five seed averages. The 40-step macro error is the prespecified primary endpoint.

Table 18: Fresh-cohort ablations with completed interaction coordinates on weighted exchange–advection–diffusion (seeds 200–204). All entries are mean ± sample standard deviation over five seeds. Accuracy and field-total errors are percentages; invariant drift is dimensionless. All metrics are lower-is-better; bolding follows the rule of Section 4.1. All models have 6,816 real trainable parameters.

Variant	Macro 10 steps (%)	Macro 40 steps (%)	Aggregate 40 steps (%)	Field totals 40 steps (%)	Invariant drift 40 steps
BalanceNO (full)	$\mathbf{2.00 \pm 2.02}$	$\mathbf{5.46 \pm 4.27}$	$\mathbf{4.79 \pm 3.78}$	3.48 ± 3.10	$(4.56 \pm 0.59)\,10^{-9}$
No projection	$\mathbf{2.27 \pm 2.39}$	$\mathbf{6.80 \pm 5.64}$	$\mathbf{6.37 \pm 5.57}$	5.91 ± 6.30	$(9.75 \pm 18.13)\,10^{-3}$
Global projection	$\mathbf{2.20 \pm 1.41}$	$\mathbf{6.16 \pm 3.62}$	$\mathbf{6.23 \pm 3.82}$	3.95 ± 2.55	$(3.78 \pm 0.54)\,10^{-8}$
Divergence only	19.49 ± 1.73	61.02 ± 4.67	45.70 ± 3.42	128.81 ± 11.98	$\mathbf{(3.39 \pm 0.28)\,10^{-9}}$
Projection only	11.22 ± 1.36	23.70 ± 2.29	26.77 ± 3.03	$\mathbf{1.38 \pm 1.52}$	$\mathbf{(3.44 \pm 0.62)\,10^{-9}}$
No preconditioning	$\mathbf{2.48 \pm 2.66}$	18.69 ± 12.68	15.79 ± 11.08	3.81 ± 2.46	$(4.12 \pm 0.58)\,10^{-9}$

F.4 Results and interpretation

Full BalanceNO achieves the smallest mean 40-step macro error, 5.46%, compared with 6.16% for the strongest ablated control (global projection), a relative reduction of 11.4%. Against global projection, BNO has lower primary error in 3 of five paired seeds. The descriptive paired 95% Student-t interval for full-minus-control error is $[-5.112, 3.712]$ percentage points. This post-hoc comparison assumes approximately normal seed differences and is not multiplicity-adjusted; it is not a claim of statistically established superiority.

Table 19: Retained initial study using the original 2D decoder, seeds 0–4. This failed confirmation round motivated the declared amendment. The seed cohort differs from Table 18; cross-table differences are not a controlled estimate of the revision effect. Percentages are mean ± sample standard deviation over five seeds.

Variant	Macro, 40 steps (%)	Field totals (%)	Invariant drift
BalanceNO (full)	$\mathbf{4.06 \pm 3.37}$	1.75 ± 1.52	$(3.88 \pm 0.44)\,10^{-9}$
No projection	$\mathbf{3.99 \pm 2.92}$	1.97 ± 1.74	$(1.75 \pm 2.99)\,10^{-3}$
Global projection	$\mathbf{3.79 \pm 2.24}$	$\mathbf{1.42 \pm 0.77}$	$(3.07 \pm 0.40)\,10^{-8}$
Divergence only	57.12 ± 2.49	126.63 ± 7.45	$(3.50 \pm 0.30)\,10^{-9}$
Projection only	24.11 ± 2.05	$\mathbf{0.96 \pm 0.73}$	$\mathbf{(3.10 \pm 0.39)\,10^{-9}}$
No preconditioning	49.71 ± 38.72	17.10 ± 16.92	$(4.26 \pm 0.87)\,10^{-9}$

The physical invariant drift of full BNO is 4.56×10^{-9}, compared with 9.75×10^{-3} without the interaction projection. The independently generated reference has mean drift 1.08×10^{-9} after float32 storage. The divergence-only and projection-only models retain small invariant drift but have 40-step macro errors of 61.02% and 23.70%, respectively. Their poor accuracy is consistent with their incorrect additional constraints: the former cannot reproduce evolving component totals, whereas the latter cannot move the conserved density through the domain. These are architectural restrictions, not merely weaker conservation-penalty weights. The analytically closest predictions satisfying the incorrect constraints provide lower bounds: 60.97% macro error for fixed component totals and 23.25% aggregate error for a fixed pointwise conserved density. These bounds separate unavoidable representation error from finite-budget optimization error.

Removing the implemented auxiliary preconditioned path yields a macro error of 18.69%. The full-versus-global-projection comparison is the more demanding control, because both preserve the correct global invariant and share the same feature extractor. Their difference is the filtered complementary-interaction route identified above. Its result is evidence about this finite-capacity parameterization on a band-limited dataset, not a theorem that local projection is always more accurate than global correction.