PMosFM: Preconditioned Manifold Matching for One-Step Physics-Constrained Generation

Anonymous authors
Paper under double-blind review

Abstract

Physics-constrained generative models aim to generate physical fields that match a target distribution and satisfy prescribed constraints. However, enforcing these constraints often increases sampling costs through iterative corrections or training costs through residual optimization and trajectory unrolling. To address this issue, we introduce **P**reconditioned **M**anifold **o**ne-**s**tep **F**low **M**atching (**PMosFM**), a preconditioned manifold matching framework for one-step physics-constrained generation. By encoding constraints in a manifold decoder, PMosFM learns transport in intrinsic coordinates without separate residual losses or terminal residual unrolling. A geometric preconditioner rescales coordinates using the decoder-induced metric, while a regularized covariance transform approximately whitens the interpolation-state inputs. A finite-interval objective couples velocity supervision with consistency between decoded endpoints in physical space. We show that exact parameterization removes residual-induced Gauss–Newton curvature, that geometric and covariance effects separate in a local conditioning bound, and that physical flow-map error bounds endpoint distributional error. Controlled ablations examine conditioning, and experiments evaluate optimizer-update time and memory footprint. At inference, PMosFM uses one neural transport evaluation followed by physical decoding. Experiments across benchmarks show lower training and sampling time than the multi-step baselines at comparable physical and distributional fidelity. Code and datasets will be released publicly.

1 Introduction

Flow matching (FM) learns probability transport from interpolation pairs and extends to function-valued physical fields (Lipman et al., 2023; Kerrigan et al., 2024). FNO (Li et al., 2021) and DeepONet (Lu et al., 2021) approximate deterministic PDE solution operators, while physics-informed training penalizes governing equation residuals (Raissi et al., 2019). In contrast, conservation-aware models and differentiable solvers project solutions onto constrained manifolds (Hansen et al., 2023; Négiar et al., 2023). For uncertain parameters or partial observations, physics-constrained generation aims to recover distributions of physical states consistent with the prescribed constraints.

Physical information enters generative models through training losses or sampling operations. Physics-informed diffusion models penalize residuals during training (Bastek et al., 2025), while PCFT adapts pretrained flows through weak-form residuals (Tauberschmidt et al., 2026). CoCoGen and DiffusionPDE guide sampling with physical residuals (Jacobsen et al., 2025; Huang et al., 2024); D-Flow optimizes source noise through differentiable generation (Ben-Hamu et al., 2024). Projected diffusion applies constraint projections during sampling (Christopher et al., 2024).

Existing methods face different computational bottlenecks: inference-time projection vs. training-time unrolling. ECI (Cheng et al., 2025) and PCFM (Utkarsh et al., 2026) can reuse pretrained models, but constraint handling remains in the online sampling loop. PBFM (Baldan et al., 2026) avoids physics-specific correction at inference by adding residual-gradient computations and terminal unrolling during training. Studies of physics-informed learning also identify gradient imbalance and ill-conditioning as optimization difficulties (Krishnapriyan et al., 2021; Wang et al., 2021; Cao & Zhang, 2025). Training-time enforcement shifts physics computation offline, while residual optimization and iterative learned transport remain. This raises a natural question: ***can physical constraint satisfaction and transport be unified to reduce both training and sampling costs?***

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.

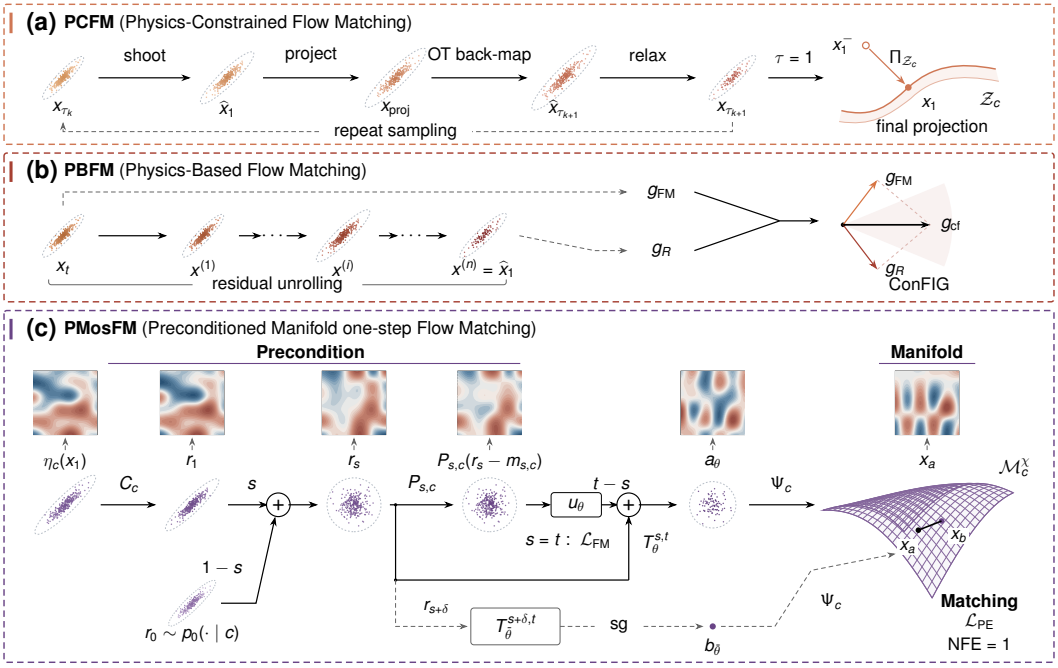


Figure 1: Overview of PMosFM and representative physics-constrained methods. (a) PCFM projects a pre-correction endpoint x_1^- onto the feasible set $\mathcal{Z}_c$ using $\Pi_{\mathcal{Z}_c}$. (b) PBFM combines flow and residual gradients g_{FM} and g_R into the ConFIG direction g_{cf}. (c) PMosFM applies precondition–manifold–matching, using $T_\theta^{0,1}$ for one-step transport and Ψ_c for physical decoding.

Recent work reduces iterative generative transport through path straightening (Liu et al., 2023), consistency (Song et al., 2023), or direct finite-interval learning (Han et al., 2026). Physics-informed distillation further enables one-step physics-constrained generation (Zhang et al., 2026). However, one-step physical generation must address not only transport depth, but also the representation of feasible states and the metric in which finite-interval errors are measured. Meanwhile, manifold generative models extend transport to non-Euclidean state spaces (Mathieu & Nickel, 2020; De Bortoli et al., 2022; Zhong et al., 2026), while preconditioned flow matching improves optimization under anisotropic interpolation statistics (Ahamed et al., 2026). Feasible coordinates expose two distinct conditioning factors: decoder-induced metric distortion and interpolation-state covariance.

In this work, we introduce **P**reconditioned **M**anifold **o**ne-**s**tep **F**low **M**atching (**PMosFM**) to address these challenges via the proposed *precondition–manifold–matching* strategy. For an exact residual parameterization on the valid domain, PMosFM removes a separate residual penalty and terminal residual unrolling for the encoded constraints. Manifold and interpolation-state preconditioners condition feasible coordinates and network inputs. A finite-interval objective couples velocity supervision with endpoint consistency in physical space. Together, these components define a one-step generator evaluated by one finite-interval neural map followed by physical decoding (Fig. 1).

We summarize the main contributions as follows:

- **One-step physics-constrained generation.** We formulate finite-interval transport in feasible coordinates with one-step evaluation followed by physical decoding. For an exact parameterization on the valid domain, the representation enforces the encoded residual directly, replacing separate training penalties and repeated unrolling.
- **Precondition–manifold–matching strategy.** We formulate physical endpoint matching with geometric and input-covariance preconditioning. Velocity supervision at $s = t$ learns the interpolation field, while endpoint consistency at $s < t$ trains finite-interval transport in feasible coordinates.
- **Feasibility, conditioning, and transport.** We show that exact residual coordinates remove the Gauss–Newton curvature contributed by the encoded constraint, while geometric and

Table 1: Comparison between PMOSFM and other physics-constrained generative models.

Method	Physics at training	Hard constraints	Gradient-free inference	Complex constraints	No manual physics–fidelity balancing	One step sampling	Residual-factorized physics training	Spectral/Jacobian preconditioning
FFM [20]	×	✓	✓	×	×	×	×	×
FM-OT [24]	×	×	✓	×	×	×	×	×
CoCoGen [19]	×	✓	✓	✓	×	×	×	×
DiffusionPDE[18]	×	×	✓	✓	×	×	×	×
PIDM [4]	✓	×	✓	✓	×	×	×	×
D-Flow [5]	×	×	✓	✓	×	×	×	×
ECI [8]	×	✓	✓	×	×	×	×	×
PCFM [40]	×	✓	✓	✓	×	×	×	×
CCFM [23]	×	✓	✓	✓	×	×	×	×
PCFT [38]	✓	×	✓	✓	×	×	×	×
PBFM [3]	✓	×	✓	✓	×	×	×	×
PMosFM	✓	✓	✓	✓	✓	✓	✓	✓

statistical preconditioning control the conditioning of the remaining transport. We further bound endpoint distribution error by physical flow-map error and distinguish constraint satisfaction from matching the target law.

- **Efficient training and one-step sampling.** PMosFM reduces training time per update compared with training with residual unrolling and achieves faster convergence with the preconditioned residual-manifold formulation. Across five benchmarks, PMosFM outperforms the multi-step baselines in distributional accuracy and sampling speed, using a single network evaluation followed by physical decoding.

2 Related Work

Physics-constrained generative modeling. Physics enters training through residual losses or virtual observables (Bastek et al., 2025; Rixner & Koutsourelakis, 2021). PBFM couples residual optimization, conflict-free gradients, and unrolling (Baldan et al., 2026; Liu et al., 2025), whereas PCFT applies weak-form fine-tuning (Tauberschmidt et al., 2026). Related approaches include residual-gradient conditioning (Shu et al., 2023), diffusion posterior sampling (Chung et al., 2023), and manifold-constrained diffusion (Chung et al., 2022). During sampling, CoCoGen and DiffusionPDE impose physical residuals (Jacobsen et al., 2025; Huang et al., 2024), while D-Flow optimizes source noise through differentiable generation (Ben-Hamu et al., 2024). Hard-constraint methods include ECI and PCFM (Cheng et al., 2025; Utkarsh et al., 2026), while projected diffusion (Christopher et al., 2024) and chance-constrained (Liang et al., 2025) flow matching provide alternative constraint-handling strategies.

One-step generative transport. FM learns velocity fields for finite-dimensional and function-valued data (Lipman et al., 2023; Kerrigan et al., 2024). Stochastic interpolants unify flow and diffusion objectives (Albergo et al., 2025), while minibatch optimal transport modifies source–target coupling (Tong et al., 2024). Rectified flows simplify transport paths (Liu et al., 2023); consistency and shortcut models support one-step generation (Song et al., 2023; Frans et al., 2025). SoFlow combines flow matching with solution consistency for one-step generation (Luo et al., 2026). InstaFlow distills a straightened flow (Liu et al., 2024), whereas MeanFlow learns finite-interval average velocities (Geng et al., 2025). W-Flow formulates one-step generation via distributional gradient flows (Han et al., 2026), whereas PIDDM enforces PDE constraints through post-hoc distillation for one-step generation (Zhang et al., 2026).

Manifold and preconditioning. Riemannian continuous flows, score models, and manifold ODEs incorporate prescribed non-Euclidean geometry (Mathieu & Nickel, 2020; De Bortoli et al., 2022; Lou et al., 2020). Manifold-aware FM extends vector-field regression to such spaces (Chen & Lipman, 2024), and Riemannian MeanFlow learns finite-interval transport through tangent-space alignment (Zhong et al., 2026). Preconditioned Flow Matching addresses interpolation-covariance anisotropy through invertible transformations (Ahamed et al., 2026).

Table 1 compares PMosFM with representative baselines, highlighting the preconditioned manifold formulation for one-step physics-constrained generation.

3 METHODOLOGY

3.1 PRELIMINARIES

Let $\mathcal{X}_h \subseteq \mathbb{R}^d$ be the discrete state space on mesh h, with state $x \in \mathcal{X}_h$ and d degrees of freedom. And let c collect the conditioning variables, including manifold, forcing, boundary data, observations, and physical time when applicable. We use $t \in [0, 1]$ for sampling time, distinct from the physical time when included in c. For the physical residual $R_h(\cdot, c)$, define the feasible set

$$\mathcal{Z}_c = \{x \in \mathcal{X}_h : R_h(x, c) = 0\}. \tag{1}$$

We model a regular component of $\mathcal{Z}_c$; the residual specifies feasible support and a target law specifies probability mass on it. Physical errors are measured by $\|e\|_{M_h}^2 = e^\top M_h e$, where M_h is a symmetric positive-definite quadrature matrix.

3.2 PHYSICS-BASED FLOW MATCHING

Let ν_0 be the ambient source law and $p_{\text{data}}(\cdot \mid c)$ the conditional data law. For an endpoint pair $x_0 \sim \nu_0$ and $x_1 \sim p_{\text{data}}(\cdot \mid c)$, consider ambient flow matching along

$$x_\tau = (1 - \tau)x_0 + \tau x_1, \qquad w_x = x_1 - x_0, \tag{2}$$

where $\tau \in [0, 1]$. A neural velocity field v_ϕ, with trainable parameters ϕ, is fitted to w_x using

$$\mathcal{L}_{\text{FM}}^{\text{amb}} = \mathbb{E}\left[\frac{1}{2d}\left\|v_\phi(x_\tau, \tau, c) - w_x\right\|_2^2\right], \tag{3}$$

The expectation averages over sampled endpoint pairs and interpolation times, and over conditions when training jointly across c. Ambient interpolation may leave the feasible set, so existing methods impose physical constraints during training (Baldan et al., 2026) or sampling (Utkarsh et al., 2026). Residual penalties can make endpoint regression ill-conditioned (Lemma 3 in Appendix G; mesh-refinement analysis in Appendix H.3).

3.3 PMOSFM: PRECONDITIONED MANIFOLD ONE-STEP FLOW MATCHING

PMosFM learns one-step transport on a represented feasible manifold, applying coordinate preconditioning and endpoint matching in physical space. Fig. 2 compares the methods in a schematic embedding space: **(a)** PCFM alternates full-space integration with iterative projection onto constraints; **(b)** PBFM steers full-space paths with residual gradients unrolling; and **(c)** PMosFM decodes two endpoint estimates and compares them on the represented residual manifold (inset).

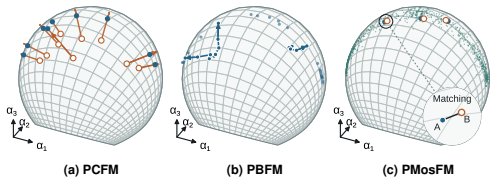


Figure 2: Manifold coordinates and endpoint matching. Schematic comparison of PCFM, PBFM, and PMosFM.

Feasible parameterization. Let $y \in \mathcal{Y}_c \subset \mathbb{R}^m$ collect the m independent coordinates on the represented feasible component. On the represented regular region, assume a differentiable local parameterization $\chi_c : \mathcal{Y}_c \to \mathcal{X}_h$ whose Jacobian has m linearly independent columns and satisfies

$$R_h(\chi_c(y), c) = 0 \qquad \text{for every } y \in \mathcal{Y}_c. \tag{4}$$

The represented residual manifold is $\mathcal{M}_c^\chi = \chi_c(\mathcal{Y}_c) \subseteq \mathcal{Z}_c$. Every decoded state satisfies the encoded constraints within the valid coordinate domain. Lemma 6 in Appendix H gives the differential characterization. The decoder and encoded-residual evaluator use the same discrete operators, boundary conditions, and sign conventions. The constant-rank and coordinate assumptions are stated in Appendix F. The encoder $\eta_c : \mathcal{M}_c^\chi \to \mathcal{Y}_c$ returns coordinates satisfying $\chi_c(\eta_c(x)) = x$ for represented data. For data outside the represented set, encoding followed by decoding produces a measurable physical-metric projection (Appendix H.6). We distinguish parameterization error from velocity-regression error.

Let $\mu_c \in \mathbb{R}^m$ be a coordinate center and $C_c \in \mathbb{R}^{m\times m}$ a fixed invertible transform that rescales physical error across coordinate directions. The coordinate and physical decoder are

$$r = C_c(y - \mu_c), \qquad \Psi_c(r) = \chi_c\big(\mu_c + C_c^{-1} r\big)\,. \tag{5}$$

The valid r-domain is $C_c(\mathcal{Y}_c - \mu_c)$, and the results below assume that interpolation states and predicted endpoints remain in this domain. For a data sample $x_1 \sim p_{\mathrm{data}}(\cdot \mid c)$, set $r_1 = C_c(\eta_c(x_1) - \mu_c)$. PMosFM pairs the encoded target r_1 with a source sample $r_0 \sim p_0(\cdot \mid c)$. The decoded target sample is $\Psi_c(r_1)$. Appendices H.6 and H.5 cover projection error and the co-area law under hard conditioning, respectively.

Manifold and preconditioning. Geometric preconditioning rescales the decoder-induced pullback metric in feasible coordinates to reduce local metric distortion, as illustrated in Fig. 3.

Let $J_\chi = D_y\chi_c$ denote the decoder Jacobian. The physical error metric in the original coordinates is $G_c(y) = J_\chi(y,c)^\top M_h J_\chi(y,c)$. Under the coordinate transformation in Eq. 5, the metric in the r coordinates becomes

$$\widetilde{G}_c(r) = C_c^{-\top} G_c(y) C_c^{-1}. \tag{6}$$

The transformation C_c rescales the pullback metric $\widetilde{G}_c$, while the endpoint loss in Eq. 12 is evaluated directly in physical space. For an interpolation pair (r_0, r_1), define

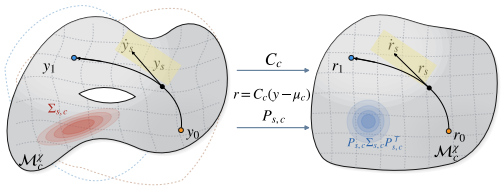


Figure 3: Preconditioned residual geometry. $y_i = \mu_c + C_c^{-1} r_i$; dots denote s-derivatives, with $\dot r_s = w$ and $\dot y_s = C_c^{-1} w$.

$$\begin{cases} r_s = (1-s)r_0 + s r_1, \\ w = r_1 - r_0, \qquad 0 \le s \le 1. \end{cases} \tag{7}$$

For each condition and generation time, define the input statistics and regularized preconditioner

$$m_{s,c} = \mathbb{E}[r_s \mid c], \quad \Sigma_{s,c} = \mathrm{Cov}(r_s \mid c), \quad P_{s,c} = (\Sigma_{s,c} + \varepsilon_P I)^{-1/2}, \tag{8}$$

where $m_{s,c}$ and $\Sigma_{s,c}$ are the conditional mean and covariance of r_s, $\varepsilon_P > 0$ regularizes the covariance, and I is the $m \times m$ identity.

The fixed C_c changes the coordinates used by Ψ_c, whereas $P_{s,c}$ rescales only the network inputs; neither changes the encoded physical target law (Appendix J).

Physical endpoint matching. PMosFM couples velocity supervision with endpoint consistency in physical space. The network u_θ, with trainable parameters θ, predicts an m-dimensional displacement rate in the r coordinates from preconditioned inputs. For $0 \le s \le t \le 1$, define the map from generation time s to t by

$$T_\theta^{s,t}(r,c) = r + (t-s)u_\theta(P_{s,c}(r - m_{s,c}), s, t, c)\,. \tag{9}$$

The losses below are averaged over sampled pairs (r_0, r_1), times (s,t,δ) where used, and conditions c when training jointly. The $s = t$ branch retains ordinary flow matching:

$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E}\left[\frac{1}{2m}\,\|u_\theta(P_{s,c}(r_s - m_{s,c}), s, s, c) - w\|_2^2\right]. \tag{10}$$

This term supervises the velocity of the interpolation path and retains the flow-matching anchor. For the $s < t$ endpoint-consistency branch, let δ be the separation between the two starting times, with $0 < \delta < t - s$. Set $r_{s+\delta} = r_s + \delta w$. The two endpoint estimates are

$$a_\theta = T_\theta^{s,t}(r_s, c), \qquad b_{\bar\theta} = T_{\bar\theta}^{s+\delta,t}(r_{s+\delta}, c), \tag{11}$$

where $\bar\theta = \mathrm{sg}(\theta)$. The stopped-gradient branch takes a current-parameter copy; gradients are stopped through this branch. Because both branches predict the endpoint at time t, PMosFM compares their decoded physical states:

$$\mathcal{L}_{\mathrm{PE}} = \mathbb{E}\left[\frac{\|\Psi_c(a_\theta) - \mathrm{sg}[\Psi_c(b_{\bar\theta})]\|_{M_h}^2}{2m\,\delta(t-s)}\right]. \tag{12}$$

In Algorithm 1, lowercase ℓ_{FM} and ℓ_{PE} denote minibatch estimates of the corresponding population losses. Within the valid coordinate domain, the decoded endpoints satisfy the encoded residual and are compared in physical space using M_h.

The complete loss is

$$\mathcal{L}_{\mathrm{PMosFM}} = \gamma \mathcal{L}_{\mathrm{FM}} + (1-\gamma)\mathcal{L}_{\mathrm{PE}}. \tag{13}$$

where the weight $\gamma \in [0,1]$ balances velocity supervision and physical-endpoint consistency.

Algorithm 1 summarizes the core PMosFM training and one-step sampling procedure. Algorithms 2 and 3 in Appendix B provide the complete training and sampling procedures.

Feasibility and conditioning. PMosFM evaluates the neural transport map, $\widehat{r}_1 = T_\theta^{0,1}(r_0, c)$, and applies the explicit decoder $\widehat{x}_1 = \Psi_c(\widehat{r}_1)$.

Proposition 1 (Exact feasibility)**.** *For fixed c and $\widehat{r}_1$ in the coordinate domain, Eq. 4 gives*

$$R_h(\widehat{x}_1, c) = 0. \tag{14}$$

The encoded residual vanishes on the valid domain, contributing no Gauss–Newton curvature (see Appendix H.2).

Algorithm 1 PMosFM: training and sampling.

Require: Data $p_{\mathrm{data}}(x \mid c)$, source $p_0(r \mid c)$, model u_θ, encoder–decoder (η_c, χ_c), metric M_h, γ.

Estimate $(\mu_c, C_c, m_{s,c}, P_{s,c})$ ▷ Precondition
while not converged **do**
  Select c; draw (s, t, δ)
  $x_1 \sim p_{\mathrm{data}}(\cdot \mid c)$, $r_0 \sim p_0(\cdot \mid c)$
  $r_1 \leftarrow C_c(\eta_c(x_1) - \mu_c)$ ▷ Feasible manifold
  $r_s \leftarrow (1-s)r_0 + s r_1$, $w \leftarrow r_1 - r_0$
  $r_{s+\delta} \leftarrow r_s + \delta w$
  Evaluate ℓ_{FM} in Eq. 10 ▷ Velocity anchor
  $\bar{\theta} \leftarrow \mathrm{sg}(\theta)$
  $a_\theta \leftarrow T_\theta^{s,t}(r_s, c)$, $b_{\bar{\theta}} \leftarrow T_{\bar{\theta}}^{s+\delta,t}(r_{s+\delta}, c)$
  Evaluate ℓ_{PE} in Eq. 12 ▷ Manifold matching
  $\theta \leftarrow \mathrm{Update}[\gamma \ell_{\mathrm{FM}} + (1-\gamma)\ell_{\mathrm{PE}}]$
end while
$r_0 \sim p_0(\cdot \mid c)$, $\widehat{r}_1 \leftarrow T_\theta^{0,1}(r_0, c)$ ▷ One-step
$\widehat{x}_1 \leftarrow \Psi_c(\widehat{r}_1)$ ▷ Physical decode
return $\widehat{x}_1$

Consider a local velocity model linear in the preconditioned inputs, with coefficient matrix $A \in \mathbb{R}^{m \times m}$. Let H_A^{GN} denote the Gauss–Newton matrix of the decoded-endpoint loss with respect to A, at fixed $s < t$, c, δ, and stopped-gradient target. Assume that the input covariance $P_{s,c}\Sigma_{s,c}P_{s,c}^\top$ is positive definite and that the eigenvalues of $\widetilde{G}_c$ lie in $[\alpha, \beta]$ throughout the evaluated endpoint region, with $0 < \alpha \le \beta$. Proposition 13 in Appendix J gives

$$\kappa(H_A^{\mathrm{GN}}) \le \frac{\beta}{\alpha}\, \kappa\big(P_{s,c}\Sigma_{s,c}P_{s,c}^\top\big), \tag{15}$$

where κ denotes the condition number. The bound separates decoder-geometry and input-covariance effects; the appendices treat approximate decoders, endpoint error, and multiple local charts.

4 Experiments

4.1 Experimental setup

Datasets. We evaluate PMosFM on Darcy flow, dynamic stall, and Kolmogorov-flow generation and reconstruction (Baldan et al., 2026), together with Burgers (Utkarsh et al., 2026). Turbulence forecasting and reconstruction use the existing turbulence benchmark (Oommen et al., 2026). We also evaluate a constrained-PDE reference suite with linear and nonlinear constraints. Dataset and evaluator details are provided in Appendices D.2 and C.

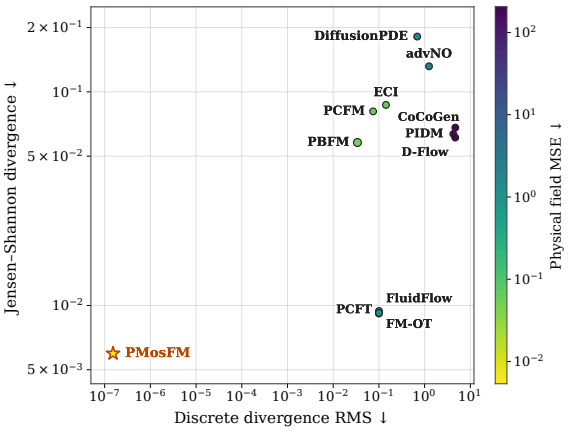


Figure 4: Physical and sampling quality comparison. Lower divergence and MSE are better.

Baselines. We compare PMosFM with representative baselines that enforce physics constraints during training or sampling. Training-time methods include PBFM (Baldan et al., 2026); sampling-time methods include ECI (Cheng et al., 2025), PCFM (Utkarsh et al., 2026), D-Flow (Ben-Hamu et al., 2024), CoCoGen (Jacobsen et al., 2025), and DiffusionPDE (Huang et al., 2024). Additional comparisons include PIDM (Bastek et al., 2025) and FM-OT (Lipman et al., 2023). Results taken from prior work are distinguished from locally obtained results.

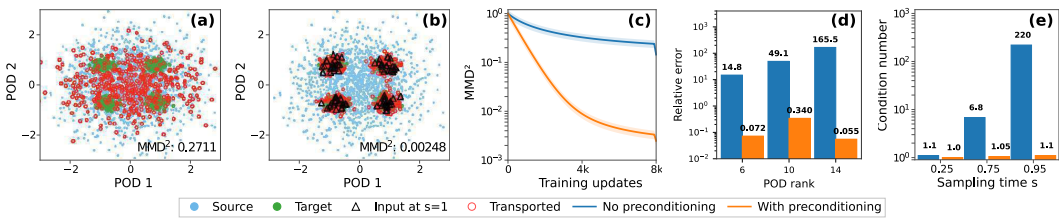


Figure 5: Preconditioning of feasible transport. (a) No preconditioning map. (b) With preconditioning map. (c) MMD^2 convergence. (d) Low-variance mode error. (e) Interpolation conditioning.

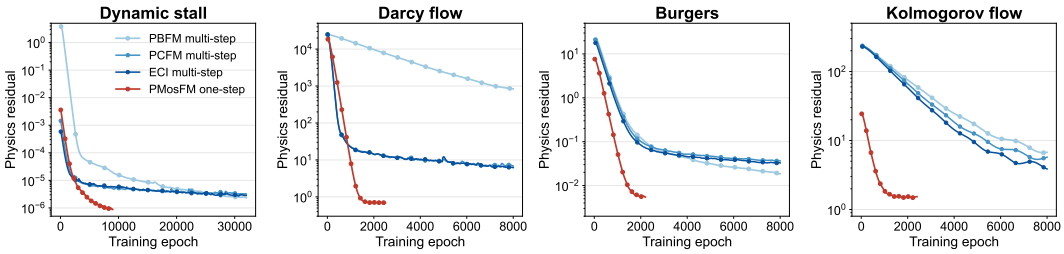


Figure 6: Training-residual convergence across physics-constrained benchmarks.

4.2 Training and Sampling Efficiency

Training cost accounting. For each task, a common physical–distributional validation criterion defines the target quality. Training cost includes encoding and preconditioner calibration, prerequisite training, and optimizer updates up to the first checkpoint meeting this criterion; validation overhead is reported separately unless full elapsed time is measured. Table 2 shows lower optimizer-update time and peak memory for PMosFM than for every evaluated PBFM unrolling depth across all five tasks. For PBFM, both quantities generally increase with unrolling depth (Baldan et al., 2026). PMosFM removes residual unrolling, while the endpoint branches and physical decoder remain part of the training cost. The training trajectories in Fig. 6 show earlier residual reduction for PMosFM, consistent with the convergence benefit sought by preconditioned manifold matching. The optimization effects of preconditioning are evaluated in Section 4.5, with training efficiency detailed in Appendix D.1.

Table 2: Optimizer-update time (s) and peak CUDA memory (GB) across unrolling configurations.

Benchmark		PBFM 1 step	PBFM 2 steps	PBFM 3 steps	PBFM 4 steps	PBFM Sum	PMosFM† 0 unrolling
Darcy	[s]	0.067	0.098	0.128	0.159	0.452	**0.019**
	[GB]	12.6	23.3	33.7	44.2	114	**0.128**
Kolmogorov	[s]	0.194	0.302	0.410	0.518	1.42	**0.021**
	[GB]	4.52	6.91	9.56	12.2	33.2	**0.143**
Dynamic Stall	[s]	0.081	0.118	0.155	0.190	0.544	**0.056**
	[GB]	4.41	7.18	9.90	12.6	34.1	**0.258**
Turbulence reconstruction	[s]	1.38	1.51	1.47	1.63	5.99	**0.930**
	[GB]	13.3	19.1	24.8	30.6	87.8	**3.50**
Turbulence forecasting	[s]	0.038	0.080	0.098	0.099	0.315	**0.021**
	[GB]	1.53	2.30	3.07	3.83	10.7	**0.584**

Sampling quality and inference cost. We next compare sampling quality and inference time under each benchmark's evaluator. Fig. 4 shows physical consistency, distributional accuracy, and field error across the evaluated samplers. Further, Table 3 compares physical and distributional errors, learned-network evaluations, and inference time across five benchmarks. Baselines rely on multi-step learned transport, requiring 20–200 network evaluations, whereas PMosFM performs one-step evaluation followed by physical decoding. In particular, PCFM enforces physical constraints during sampling and has higher inference time than PMosFM in Table 3. PMosFM also achieves competitive physical and distributional accuracy across the benchmarks in Table 3. Thus, PMosFM reduces sampling cost while maintaining competitive physical and distributional accuracy. The sampling time includes physical decoding; decoder tolerances and costs are detailed in Appendix O.2.

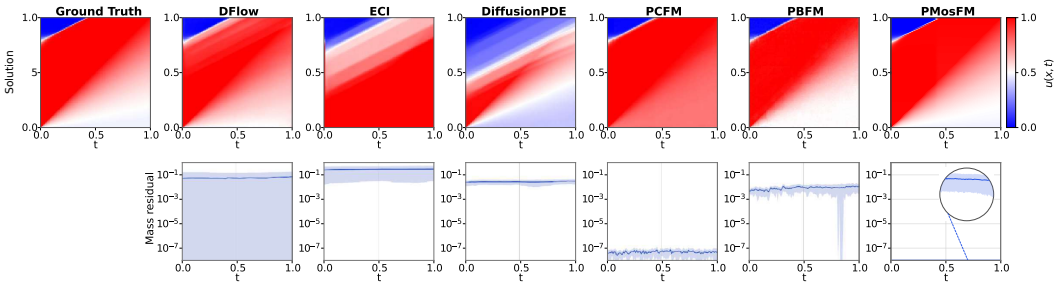


Figure 7: Generated fields and conservation residuals across methods. The top row compares generated solutions, while the bottom row shows the conservation-residual trajectories.

4.3 Physics-constrained generation

Constraint satisfaction and field fidelity. Across the physics-constrained benchmarks, PMosFM preserves the principal solution structure while keeping the encoded conservation residual near numerical precision, whereas several baselines retain localized violations around dominant transition regions (see Fig. 7). The quantitative results further show competitive distributional accuracy alongside constraint satisfaction (see Table 9 in Appendix D.3), indicating that improved physical consistency is achieved while retaining competitive distributional accuracy.

Held-out turbulence diagnostics. On held-out turbulence, PMosFM reproduces the characteristic tear-drop topology and non-Gaussian tails of the velocity-gradient statistics, while the baselines underrepresent extreme-strain events (see Fig. 13 in Appendix D.8). These statistics probe small-scale intermittency beyond the encoded constraint residual, showing that PMosFM retains fine-scale turbulence structure in one-step generation.

Table 3: Sampling quality and inference efficiency across benchmarks. RE: physical residual error; WD: Wasserstein distance; JS: Jensen–Shannon divergence; NFE: learned network evaluations; IT: inference time.

Benchmark	Metric	FM-OT	CoCoGen	PIDM	DiffusionPDE	D-Flow	ECI	PCFM	PBFM	PMosFM
Dynamic Stall	RE $\times 10^6 \downarrow$	11.0	376	2.96×10^3	12.2	11.3	40.7	0.143	0.339	**0.0421**
	WD $\times 10^4 \downarrow$	2.71	132	179	2.51	3.48	11.1	4.01	1.81	**0.842**
	JS $\times 10^2 \downarrow$	0.983	1.69	4.14	1.03	1.01	0.0570	7.46	0.680	**0.0368**
	NFE $\downarrow$	20	100	100	20	20	200	20	20	**1**
	IT [s] $\downarrow$	0.0598	0.184	0.0866	0.172	0.139	0.432	3.91	0.0605	**0.00338**
Darcy Flow	RE $\downarrow$	4.16	1.32	0.0220	3.39	2.29	3.05	4.19×10^3	0.838	**0.120**
	WD $\times 10^2 \downarrow$	0.0590	0.249	3.10	0.0890	0.147	2.89	3.74	0.138	**0.0490**
	JS $\times 10^1 \downarrow$	0.131	0.360	3.18	0.139	0.237	2.82	0.199	0.256	**0.0139**
	NFE $\downarrow$	20	100	100	20	20	20	20	20	**1**
	IT [s] $\downarrow$	0.100	7.40	2.05	0.590	3.13	0.122	1.33	0.101	**0.0178**
Burgers	RE $\times 10^1 \downarrow$	4.30	404	1.57×10^3	3.66	3.62×10^3	0.818	0.473	0.307	**0.0435**
	WD $\times 10^2 \downarrow$	4.85	36.0	248	0.685	638	2.55	2.73	4.02	**0.129**
	JS $\times 10^2 \downarrow$	1.92	12.3	31.5	0.184	52.4	0.809	0.884	1.43	**0.0320**
	NFE $\downarrow$	20	20	20	20	20	20	20	20	**1**
	IT [s] $\downarrow$	0.0349	0.0354	0.0354	0.227	0.0354	0.0902	0.134	0.0913	**0.00237**
Kolmogorov Flow	RE $\times 10^1 \downarrow$	2.31	13.7	14.3	1.93	6.74	0.000	1.53	1.36	**0.0842**
	WD $\times 10^1 \downarrow$	2.12	3.11	2.86	3.70	1.01	0.685	0.665	1.22	**0.382**
	JS $\times 10^2 \downarrow$	12.5	29.1	29.2	19.4	16.8	7.10	6.87	7.44	**1.84**
	NFE $\downarrow$	20	100	100	20	20	200	200	20	**1**
	IT [s] $\downarrow$	0.0988	0.0449	0.0506	0.268	6.43	0.385	0.792	0.0990	**0.00375**
Turbulence Flow Reconstruction	RE $\times 10^1 \downarrow$	5.28	1.52	12.1	0.767	9.52	0.521	0.521	0.192	**0.0258**
	WD $\times 10^2 \downarrow$	2.47	1.35	3.68	1.01	2.85	1.54	1.49	1.19	**0.684**
	JS $\times 10^2 \downarrow$	3.82	1.90	5.22	1.25	4.13	2.11	1.97	1.42	**0.362**
	NFE $\downarrow$	20	20	20	20	20	20	20	20	**1**
	IT [s] $\downarrow$	0.239	0.239	0.239	0.222	0.239	0.0301	0.0377	0.0268	**0.00370**

4.4 Underdetermined sparse reconstruction

Physical constraints restrict the feasible set but leave sparse reconstruction underdetermined. Fig. 8 compares PMosFM and the baselines under identical observations and masks on a held-out state. As missingness increases, the baselines lose spatial organization; with 99% missing observations, they fail to recover the dominant structure of the reference kinetic-energy field, whereas PMosFM retains

Table 4: One-step-core ablation within PMosFM. Energy distance (ED; ↓) is reported; lower is better. KF denotes Kolmogorov flow. Bold denotes the lowest available value in each benchmark row.

Benchmark	PMosFM-MeanFlow	PMosFM-iMF	PMosFM-W-Flow	PMosFM-Drifting	PMosFM-SoFlow	PMosFM-Shortcut	**PMosFM-native**
Burgers	0.0758	0.0765	0.0518	0.0713	0.0714	0.0849	**0.0374**
Dynamic Stall	0.1403	0.1315	0.0594	0.0442	0.1352	0.0455	**0.0036**
Darcy Flow	0.0277	0.0341	0.0163	0.0233	0.0317	0.0158	**0.0043**
KF Reconstruction	0.0144	0.0145	0.0574	0.0476	0.0249	0.0104	**0.0037**
KF Generation	2.0645	2.1278	1.6020	2.1451	2.0887	1.9672	**1.5922**
Turbulence Reconstruction	1.0776	1.0836	1.0829	1.0829	1.0727	**0.9423**	0.9632
Turbulence Forecasting	1.5448	1.4719	0.8767	0.8658	1.5019	0.9752	**0.7507**

an identifiable large-scale pattern. At 99% missingness, PMosFM better preserves low-wavenumber energy and velocity-gradient PDF tails than the baselines. This test evaluates whether generated fields retain spatial and multiscale structure in unobserved regions beyond constraint satisfaction.

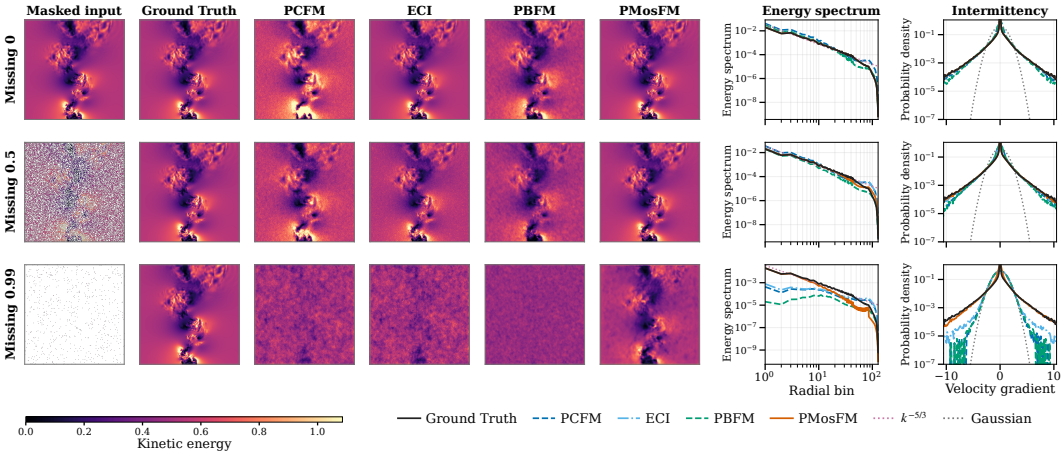


Figure 8: Structural recovery under sparse observations. PMosFM retains large-scale structure under masking across field, spectral, and velocity-gradient diagnostics.

4.5 Ablation Study

Ablation experiments isolate the effects of manifold parameterization, metric preconditioning, and endpoint regression. Fig. 5 shows that preconditioning improves distribution matching, reduces errors in low-variance modes, and stabilizes the interpolation covariance seen by the network. Together, these diagnostics connect the improved training behavior to the statistical conditioning targeted by PMosFM. Table 4 benchmarks different one-step velocity regression formulations within PMosFM using energy distance (ED). PMosFM gives the lowest ED on most benchmarks, whereas Shortcut is stronger on turbulence reconstruction. The variation suggests that the preferred one-step objective depends on the physical task. Additional ablation results are reported in Appendix E.

5 Conclusion

In this work, we present PMosFM, a precondition–manifold–matching framework for one-step physics-constrained generation that separates encoded feasibility from learned transport. Feasible coordinates remove encoded residual-normal directions, while geometric and interpolation-state preconditioning condition the remaining transport, and finite-interval matching compares decoded endpoints in physical space. Across the evaluated PDE settings, PMosFM achieves competitive physical and distributional fidelity with near-zero encoded residuals; controlled studies show better conditioning and faster distributional convergence with preconditioning. PMosFM also reduces measured optimizer-update time relative to PBFM, and one-step sampling with physical decoding lowers inference time compared with the evaluated multi-step baselines.

AI USE STATEMENT

Generative AI tools were used to assist with manuscript organization, language editing, revision, and figure preparation. All AI-assisted content was reviewed and revised by the authors; experimental measurements were produced by the reported computational pipelines, and the mathematical derivations, experimental results, references were independently verified. The authors take full responsibility for the final content of this work, including all AI-assisted text and artifacts.

REFERENCES

Shadab Ahamed, Eshed Gal, Md Shahriar Rahim Siddiqui, Simon Ghyselincks, Moshe Eliasof, and Eldad Haber. Preconditioned flow matching. *arXiv preprint arXiv:2603.02337*, 2026.

Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

Giacomo Baldan, Qiang Liu, Alberto Guardone, and Nils Thuerey. Physics vs distributions: Pareto optimal flow matching with physics constraints. In *International Conference on Learning Representations*, 2026.

Jan-Hendrik Bastek, WaiChing Sun, and Dennis Kochmann. Physics-informed diffusion models. In *International Conference on Learning Representations*, 2025.

Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-Flow: Differentiating through flows for controlled generation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 3462–3483. PMLR, 2024.

Wenbo Cao and Weiwei Zhang. An analysis and solution of ill-conditioning in physics-informed neural networks. *Journal of Computational Physics*, 520:113494, 2025.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations*, 2024.

Chaoran Cheng, Boran Han, Danielle C. Maddix, Abdul Fatir Ansari, Andrew Stuart, Michael W. Mahoney, and Bernie Wang. Gradient-free generation for hard-constrained systems. In *International Conference on Learning Representations*, 2025.

Jacob K. Christopher, Stephen Baek, and Ferdinando Fioretto. Constrained synthesis with projected diffusion models. In *Advances in Neural Information Processing Systems*, volume 37, pp. 89307–89333, 2024.

Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. In *Advances in Neural Information Processing Systems*, volume 35, pp. 25683–25696, 2022.

Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023.

Valentin De Bortoli, Émile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. In *Advances in Neural Information Processing Systems*, volume 35, pp. 2406–2422, 2022.

Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One step diffusion via shortcut models. In *International Conference on Learning Representations*, 2025.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J. Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Jiaqi Han, Puheng Li, Qiushan Guo, Renyuan Xu, Stefano Ermon, and Emmanuel J. Candès. One-step generative modeling via wasserstein gradient flows. *arXiv preprint arXiv:2605.11755*, 2026.

Derek Hansen, Danielle C. Maddix, Shima Alizadeh, Gaurav Gupta, and Michael W. Mahoney. Learning physical models that can respect conservation laws. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 12469–12510. PMLR, 2023.

Maximilian Herde, Bogdan Raonić, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel de Bézenac, and Siddhartha Mishra. Poseidon: Efficient foundation models for PDEs. In *Advances in Neural Information Processing Systems*, volume 37, pp. 72525–72624, 2024.

Jiahe Huang, Guandao Yang, Zichen Wang, and Jeong Joon Park. DiffusionPDE: Generative PDE-solving under partial observation. In *Advances in Neural Information Processing Systems*, volume 37, pp. 130291–130323, 2024.

Christian Jacobsen, Yilin Zhuang, and Karthik Duraisamy. CoCoGen: Physically consistent and conditioned score-based generative models for forward and inverse problems. *SIAM Journal on Scientific Computing*, 47(2):C399–C425, 2025.

Gavin Kerrigan, Giosue Migliorini, and Padhraic Smyth. Functional flow matching. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 3934–3942. PMLR, 2024.

Aditi Krishnapriyan, Amir Gholami, Shandian Zhe, Robert Kirby, and Michael W. Mahoney. Characterizing possible failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems*, volume 34, pp. 26548–26560, 2021.

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*, 2021.

Jinhao Liang, Yixuan Sun, Anirban Samaddar, Sandeep Madireddy, and Ferdinando Fioretto. Chance-constrained flow matching for high-fidelity constraint-aware generation. *arXiv preprint arXiv:2509.25157*, 2025.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Qiang Liu, Mengyu Chu, and Nils Thuerey. ConFIG: Towards conflict-free training of physics-informed neural networks. In *International Conference on Learning Representations*, 2025.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023.

Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, and Qiang Liu. InstaFlow: One step is enough for high-quality diffusion-based text-to-image generation. In *International Conference on Learning Representations*, 2024.

Aaron Lou, Derek Lim, Isay Katsman, Leo Huang, Qingxuan Jiang, Ser-Nam Lim, and Christopher M. De Sa. Neural manifold ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 33, pp. 17548–17558, 2020.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.

Tianze Luo, Haotian Yuan, and Zhuang Liu. SoFlow: Solution flow models for one-step generative modeling. In *International Conference on Learning Representations*, 2026.

Emile Mathieu and Maximilian Nickel. Riemannian continuous normalizing flows. In *Advances in Neural Information Processing Systems*, volume 33, pp. 2503–2515, 2020.

Geoffrey Négiar, Michael W. Mahoney, and Aditi S. Krishnapriyan. Learning differentiable solvers for systems with hard constraints. In *International Conference on Learning Representations*, 2023.

Vivek Oommen, Siavash Khodakarami, Aniruddha Bora, Zhicheng Wang, and George Em Karniadakis. Learning turbulent flows with generative models for super resolution and sparse flow reconstruction. *Nature Communications*, 17:3707, 2026.

Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Maximilian Rixner and Phaedon-Stelios Koutsourelakis. A probabilistic generative model for semi-supervised training of coarse-grained surrogates and enforcing physical constraints through virtual observables. *Journal of Computational Physics*, 434:110218, 2021.

Dule Shu, Zijie Li, and Amir Barati Farimani. A physics-informed diffusion model for high-fidelity flow field reconstruction. *Journal of Computational Physics*, 478:111972, 2023.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 32211–32252. PMLR, 2023.

Jan Tauberschmidt, Sophie Fellenz, Sebastian J. Vollmer, and Andrew B. Duncan. Physics-constrained fine-tuning of flow-matching models for generation and inverse problems. In *International Conference on Learning Representations*, 2026.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024.

Utkarsh Utkarsh, Pengfei Cai, Alan Edelman, Rafael Gomez-Bombarelli, and Christopher Rackauckas. Physics-constrained flow matching: Sampling generative models with hard constraints. *Advances in Neural Information Processing Systems*, 38:160217–160252, 2026.

Sifan Wang, Yujun Teng, and Paris Perdikaris. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43(5):A3055–A3081, 2021.

Jian Xu, Yanning Wu, Delu Zeng, John Paisley, and Qibin Zhao. The right measure for physics-constrained generation: A co-area correction for posterior-consistent pde inverse problems. *arXiv preprint arXiv:2606.04804*, 2026.

Yi Zhang, Peng Wang, and Difan Zou. Physics-informed distillation of diffusion models for PDE-constrained generation. In *Forty-third International Conference on Machine Learning*, 2026.

Zichen Zhong, Haoliang Sun, Yukun Zhao, Yongshun Gong, and Yilong Yin. Riemannian MeanFlow for one-step generation on manifolds. In *International Conference on Machine Learning*, 2026.

A NOTATION

This table summarizes the notation adopted throughout the paper and appendix. Symbols are grouped by residual manifold, preconditioning, and finite-interval matching.

Table 5: Summary of the notation adopted throughout the paper and appendix.

Symbol	Name	Description
$(x, \mathcal{X}_h, c, h)$	state space, condition, mesh	$x \in \mathcal{X}_h \subseteq \mathbb{R}^d$ is a discretized physical state; c is the condition; h is the spatial discretization.
$R_h(x, c) \in \mathbb{R}^{n_R}$	discrete residual	Residual that evaluates the hard physical constraint; n_R is the number of residual components.
$(\mathcal{Z}_c, \mathcal{M}_c^\chi)$	residual-zero set	$\mathcal{Z}_c = \{x : R_h(x, c) = 0\}$; $\mathcal{M}_c^\chi = \chi_c(\mathcal{Y}_c) \subseteq \mathcal{Z}_c$ is the component represented by the model.
$(y, \mathcal{Y}_c, m)$	intrinsic coordinates	$y \in \mathcal{Y}_c \subset \mathbb{R}^m$ represents the independent degrees of freedom; $m = d - q$ under the regular-rank assumption.
(χ_c, η_c)	constraint-satisfying parameterization and encoder	$R_h \circ \chi_c \equiv 0$; η_c encodes feasible data or provides a measurable selection.
(J_χ, J_R)	coordinate and residual Jacobians	$J_\chi = D_y \chi_c$, $J_R = D_x R_h$ satisfy $J_R J_\chi = 0$; regular-rank conditions identify $\operatorname{range} J_\chi$ with the tangent space.
(M_h, M_x)	physical metric	Fixed positive-definite physical or quadrature metric; M_x abbreviates M_h when h is implicit.
$G_c(y)$	pullback metric	Physical metric in intrinsic coordinates: $G_c(y) = J_\chi(y, c)^\top M_h J_\chi(y, c)$.
(C_c, μ_c, Ψ_c)	geometric setting	$r = C_c(y - \mu_c)$ and $\Psi_c(r) = \chi_c(\mu_c + C_c^{-1} r)$.
(r_0, r_1, r_s, w)	interpolation pair	$r_s = (1 - s)r_0 + s r_1$ and $w = r_1 - r_0$.
$(m_{s,c}, \Sigma_{s,c}, P_{s,c})$	input preconditioner	Mean, covariance, and $P_{s,c} = (\Sigma_{s,c} + \varepsilon_P I)^{-1/2}$.
$(T_\theta^{s,t}, u_\theta)$	finite-interval map and velocity	$T_\theta^{s,t}(r, c) = r + (t - s) u_\theta(P_{s,c}(r - m_{s,c}), s, t, c)$.
$(a_\theta, b_{\bar\theta})$	endpoint estimates	Endpoint estimates from nearby interpolation times, with gradients stopped through the second branch.
$(\mathcal{L}_{\mathrm{FM}}, \mathcal{L}_{\mathrm{PE}})$	matching losses	Diagonal velocity FM loss and off-diagonal decoded-endpoint loss.
(e_θ, d_θ)	map defects	Coordinate flow-map defect and the corresponding decoded physical defect.
$(\gamma, \delta, \varepsilon_P)$	matching parameters	Branch weight, finite-difference interval, and covariance regularizer.
(a, J_G)	differentiation variable and endpoint Jacobian	a is an input or parameter; $J_G = D_a G_\phi$ is the corresponding endpoint Jacobian.
$H^R_{\mathrm{GN},a}$	residual Gauss–Newton block	$J_G^\top J_R^\top W_R J_R J_G$; exact residual factorization makes this block vanish for the encoded residual after composition with the decoder.
$(\varepsilon_{\mathrm{chart}}, \delta_{\mathrm{chart}})$	approximation error bounds	Bounds on residual magnitude and pulled-back residual differential.
$(\sigma^+_{\min}, \kappa_+)$	active spectral quantities	Smallest nonzero singular value and condition number on the active subspace.

B DETAILED TRAINING AND SAMPLING PROCEDURES

The main-text algorithm exposes the computation that distinguishes PMosFM. This appendix separates calibration and training from one-step physical sampling.

Algorithm 2 Full training procedure of PMosFM

Require: Data $p_{\mathrm{data}}(x \mid c)$; source $p_0(r \mid c)$
Model u_θ; encoder–decoder (η_c, χ_c)
Metric M_h; γ; ε_P
Ensure: Trained θ and frozen $(\mu_c, C_c, m_{s,c}, P_{s,c})$
Estimate (μ_c, C_c) from the training data
Using $r_1 = C_c(\eta_c(x_1) - \mu_c)$ and $r_s = (1-s)r_0 + sr_1$, estimate $m_{s,c} \leftarrow \mathbb{E}[r_s \mid c]$, $\Sigma_{s,c} \leftarrow \mathrm{Cov}(r_s \mid c)$, $P_{s,c} \leftarrow (\Sigma_{s,c} + \varepsilon_P I)^{-1/2}$
Fix $(\mu_c, C_c, m_{s,c}, P_{s,c})$
while not converged **do**
  Select c; draw $0 \le s < t \le 1$ and $0 < \delta < t - s$
  $x_1 \sim p_{\mathrm{data}}(\cdot \mid c)$, $r_0 \sim p_0(\cdot \mid c)$
  $r_1 \leftarrow C_c(\eta_c(x_1) - \mu_c)$
  $r_s \leftarrow (1-s)r_0 + sr_1$, $w \leftarrow r_1 - r_0$
  $r_{s+\delta} \leftarrow r_s + \delta w$
  $\ell_{\mathrm{FM}} \leftarrow \dfrac{1}{2m} \left\| u_\theta(P_{s,c}(r_s - m_{s,c}), s, s, c) - w \right\|_2^2$
  $\bar{\theta} \leftarrow \mathrm{sg}(\theta)$
  $a_\theta \leftarrow T_\theta^{s,t}(r_s, c) = r_s + (t-s)u_\theta\big(P_{s,c}(r_s - m_{s,c}), s, t, c\big)$
  $b_{\bar{\theta}} \leftarrow T_{\bar{\theta}}^{s+\delta,t}(r_{s+\delta}, c) = r_{s+\delta} + (t-s-\delta)u_{\bar{\theta}}\big(P_{s+\delta,c}(r_{s+\delta} - m_{s+\delta,c}), s+\delta, t, c\big)$
  $\ell_{\mathrm{PE}} \leftarrow \dfrac{\|\Psi_c(a_\theta) - \mathrm{sg}[\Psi_c(b_{\bar{\theta}})]\|_{M_h}^2}{2m\,\delta(t-s)}$
  $\theta \leftarrow \mathrm{Update}[\theta,\ \gamma\ell_{\mathrm{FM}} + (1-\gamma)\ell_{\mathrm{PE}}]$
end while
return θ and $(\mu_c, C_c, m_{s,c}, P_{s,c})$

Algorithm 3 Full one-step sampling procedure of PMosFM

Require: Trained θ; source $p_0(r \mid c)$; condition c; frozen $(\mu_c, C_c, m_{0,c}, P_{0,c})$; decoder Ψ_c
Ensure: Physics-consistent sample $\widehat{x}_1$
Sample $r_0 \sim p_0(\cdot \mid c)$
Form the centered source coordinate
$$r_0 - m_{0,c}$$
Apply the frozen input preconditioner
$$P_{0,c}(r_0 - m_{0,c})$$
Evaluate the full-interval map once
$$\widehat{r}_1 \leftarrow T_\theta^{0,1}(r_0, c)$$
Expand the one-step map as
$$\widehat{r}_1 = r_0 + u_\theta(P_{0,c}(r_0 - m_{0,c}), 0, 1, c)$$
Apply the inverse geometric transform
$$\mu_c + C_c^{-1}\widehat{r}_1$$
Decode on the feasible manifold
$$\widehat{x}_1 \leftarrow \chi_c\big(\mu_c + C_c^{-1}\widehat{r}_1\big) = \Psi_c(\widehat{r}_1)$$
return $\widehat{x}_1$

The calibration quantities are estimated from the training split and then held fixed. Algorithm 2 expands the two-time objective in Eq. 13, while Algorithm 3 evaluates $T_\theta^{0,1}$ once before physical decoding. Iterations in a solver-based decoder belong to physical decoding rather than learned transport and are accounted for separately in Appendix O.2.

C Dataset Details

This section records the PDE system, data split, conditioning variables, normalization, and physical diagnostic for each benchmark. The evaluated tasks cover elliptic flow, nonlinear conservation laws, compressible aerodynamics, turbulence, and inverse problems. All normalization statistics are estimated on the training partition and remain fixed for validation and test data. Generated fields are mapped back to physical units before residual evaluation whenever inverse normalization is available. Each benchmark-specific comparison shares one split and one evaluator. When the benchmark provides the state, forcing, manifold, and boundary variables required by the governing equations, we evaluate the full prescribed residual; otherwise, we report the identifiable residual components.

C.1 Darcy Flow

Each sample contains a pressure field $p(\boldsymbol{x})$ and permeability field $K(\boldsymbol{x})$ on a two-dimensional grid. The data follow the steady Darcy system

$$\boldsymbol{u} = -K\nabla p, \qquad \nabla\cdot\boldsymbol{u} = f, \qquad \boldsymbol{u}\cdot\boldsymbol{n} = 0 \text{ on } \partial\Omega, \qquad \int_\Omega p\,\mathrm{d}\boldsymbol{x} = 0.$$

The benchmark pairs log-Gaussian permeability with steady Darcy pressure, following PBFM (Baldan et al., 2026). Both fields are standardized by channel. For common evaluation, predictions are de-normalized before applying the prescribed residual operator.

C.2 Dynamic Stall

The data are obtained from unsteady compressible two-dimensional RANS simulations of a sinusoidally pitching NACA0012 airfoil (Baldan et al., 2026). The operating condition contains the free-stream Mach number, mean angle of attack, pitching amplitude, and reduced frequency. The dataset contains 128 nominal training conditions with 32 realizations per condition and 16 held-out conditions. The six generated channels encode pressure, two tangential-velocity-gradient components, temperature, density, and signed wall shear stress on a 128×128 spatio-temporal surface grid; two additional channels provide the local surface direction. After de-normalization, the two algebraic residuals are

$$r_{\text{gas}} = p - \rho RT, \qquad r_\tau = \tau_w - \operatorname{sign}(\boldsymbol{\mu} \cdot \boldsymbol{s}) \|\boldsymbol{\mu}\|_2,$$

where $\boldsymbol{\mu} = \mu(T)(\partial_x u_s, \partial_y u_s)$, $\boldsymbol{s}$ is the local surface direction, and $\mu(T)$ is given by Sutherland's law. The manifold-conditioned implementation follows the same local direction as the external residual, so the parameterization and evaluator share a sign convention.

C.3 Burgers

The Burgers benchmark consists of trajectories at viscosity $\nu = 10^{-3}$. The first 1,000 trajectories form the training partition, indices 1,000–1,499 form validation, and the remaining trajectories form the test partition. We remove the duplicated terminal slice and standardize the resulting space–time field with training statistics. For a generated trajectory $u(t, x)$, the diagnostic is

$$r_{\text{B}} = \partial_t u + u\, \partial_x u - \nu\, \partial_{xx} u,$$

with centered differences, periodic padding in space, and replicated temporal endpoints. We also apply periodic RK4 pseudospectral evolution; because its discretization differs from the finite-difference evaluator, we report solver results separately from residual RMS.

C.4 Turbulence Mass Transfer

Each sample contains horizontal velocity, vertical velocity, pressure, and signed distance to the immersed manifold as four aligned channels. The first 900 samples form the training set, and the remainder forms the validation set with channel-wise training normalization. From these variables, we define a field-level manifold-consistency evaluator comprising discrete incompressibility, normal velocity near the zero level set, signed-distance Eikonal error, and weak pressure smoothness:

$$R_{\text{TMT}} = \left[\nabla \cdot \boldsymbol{u}, e^{-(\phi/\delta)^2} \boldsymbol{u} \cdot \boldsymbol{n}_\phi, 0.2(\|\nabla \phi\|_2 - 1), 0.02 \Delta p\right].$$

C.5 Kolmogorov Flow

The conditional dataset is generated at Reynolds numbers in $[100, 500]$ on a 128×128 periodic grid. The training set covers 32 Reynolds numbers and the held-out set 16 Reynolds numbers, with 1,024 statistically stationary velocity snapshots per condition. The two output channels are the planar velocity components, and the Reynolds number is the conditioning variable. Both the encoded residual and the independent parameterization comparison measure incompressibility,

$$r_{\text{KF}} = \partial_x u + \partial_y v,$$

after restoring channel scales. The streamfunction parameterization and Fourier evaluator use compatible periodic derivatives. Central differences reveal grid sensitivity in the residual.

C.6 Turbulence Forecasting

For turbulence forecasting, we adopt the homogeneous-isotropic-turbulence benchmark of Oommen et al. (2026). Each state contains four physical channels on a three-dimensional periodic grid. Each target is conditioned on the preceding four frames, with a one-frame forecast horizon. After the four-frame source offset, frames 0–159 form the training set, frames 160–179 form the validation set, and the remaining complete windows form the test set. Normalization follows the benchmark means and standard deviations when available. The physical consistency diagnostic is the periodic three-dimensional divergence of the first three channels,

$$r_{\text{div}} = \partial_x u + \partial_y v + \partial_z w,$$

A Fourier Helmholtz projection acts on the three velocity channels; the fourth is unchanged.

C.7 Turbulence Flow Reconstruction

This task contains four-channel, three-dimensional Gen4Turbulence fields and constructs random point masks on each frame. Frames 0–149, 150–167, and 168 onward define the training, validation, and test partitions, respectively. The condition concatenates the normalized observed field $\boldsymbol{y} = \mathbf{M} \odot \boldsymbol{x}$ and the binary mask $\mathbf{M}$; validation and test masks are fixed functions of the frame index. The common residual stacks measurement consistency with velocity divergence,

$$R_{\mathrm{rec}}(\boldsymbol{x}; \boldsymbol{y}, \mathbf{M}) = \big[\tfrac{1}{4}(\mathbf{M} \odot \boldsymbol{x} - \boldsymbol{y}),\ \nabla \cdot \boldsymbol{u}\big].$$

The one-pass decoder first performs a Fourier divergence-free projection and then overwrites observed entries. Because the observation overwrite can reintroduce divergence after the Fourier projection, we evaluate measurement consistency and incompressibility separately.

C.8 Kolmogorov-flow Generation

The KF generation benchmark consists of planar Kolmogorov-flow velocity fields (u, v), represented as trajectories or individual snapshots. Trajectory and time axes are flattened into a common sample axis, followed by a deterministic 90/10 train–validation partition and channel-wise training normalization. The diagnostic concatenates periodic divergence, a weak Laplacian penalty on vorticity, and a weak Laplacian penalty on kinetic energy:

$$R_{\mathrm{KFG}} = \big[\nabla \cdot \boldsymbol{u},\ 0.02\Delta\omega,\ 0.01\Delta(u^2 + v^2)\big]. \tag{16}$$

Divergence checks incompressibility; the other terms capture high-frequency artifacts.

C.9 Kolmogorov-flow Reconstruction

The target fields, split, and normalization are identical to Kolmogorov-flow Generation; the model is conditioned on a low-resolution observation. During training, a factor is drawn from the set $\{1, 2, 4\}$; the velocity is subsampled by that factor and bilinearly returned to the target grid before conditioning the model. The evaluation residual is the same R_{KFG} as in Eq. 16, while reconstruction error is computed against the full-resolution held-out target. Separating observation fidelity from flow diagnostics keeps reconstruction accuracy and physical consistency as distinct criteria.

D Additional Experiments

This appendix reports supplementary evaluation protocols and results for the PMosFM objective in Eq. 13. Validation data are used for model and checkpoint selection, and all normalization statistics are estimated from the training partition. PMosFM preconditioners are calibrated on the training data and held fixed for validation and test evaluation.

Measures and comparison protocol. Residual checks are applied in physical units whenever inverse normalization is available. The analytic study of the feasible set and target distribution reports residual RMS, KL divergence, and total variation against the specified co-area law. The conditioning study reports active condition numbers and the iterations required by a fixed stable-step rule. The constrained-PDE reference protocol reports MMSE, SMSE, constraint error, and Fréchet Poseidon distance; lower values indicate closer agreement with the reference distribution or constraint. Physics-constrained generation evaluations report task-specific field errors, prescribed residuals or named proxies, and distributional statistics. Timing records separate learned-network evaluations from preconditioning and physical decoding costs. Each table caption identifies the common evaluator and whether the comparison shares training budgets or matched quality.

Table 6: Summary of constraint types in the experimental suite, categorized by mathematical form and scope.

Constraint type	**Representative form**	**Linearity**
Affine IC / BC / observations	$R_h(x,c) = Ax - b = 0,$ $x = x_p + Ny,\ AN = 0$	Linear
Global conservation	$\ell^\top x - C = 0,$ conservation-preserving or mean-free coordinates	Linear
Differential compatibility	$D_x u + D_y v = 0,$ $(u,v) = (D_y\psi, -D_x\psi)$	Linear
Algebraic constitutive relation	$g(x,c) = 0,$ eliminate dependent variables	Nonlinear
Implicit PDE constraint	$F_h(z^\star(y,c), y, c) = 0,$ $z = z^\star(y,c)$	Potentially nonlinear
General nonlinear coupled constraint	$R_h(x,c) = 0,$ local parameterization or approximate decoder	Nonlinear

D.1 Comprehensive Training Cost and Memory Footprint

For a complete method, training and sampling costs are accounted for separately as

$$T_{\text{train}}(\epsilon) = T_{\text{calibration}} + N_{\text{update}}(\epsilon)\,\bar{t}_{\text{update}},$$
$$T_{\text{sample}} = T_{\text{network}} + T_{\text{precondition}} + T_{\text{decode}}.$$

where ϵ denotes a common quality criterion, $N_{\text{update}}(\epsilon)$ is the number of optimizer updates required to first meet this criterion, and $\bar{t}_{\text{update}}$ is the mean update time. $T_{\text{calibration}}$ includes estimation of the fixed preconditioner, while T_{network}, $T_{\text{precondition}}$, and T_{decode} denote learned-network evaluation, preconditioning, and physical decoding costs, respectively.

For a curriculum of PBFM unrolling depths, the optimizer-update component is computed from the actual number of updates executed at each depth:

$$T_{\text{updates}} = \sum_n U_n\,\bar{t}_n,$$

where U_n counts executed updates and $\bar{t}_n$ is their mean time at depth n. The single-update measurements in Table 2 represent separate unrolling configurations, not additive components; prerequisite training and calibration count once, and validation overhead is included only in full elapsed time.

Implementation details. Training-cost profiling used an NVIDIA RTX PRO 6000 Blackwell Server Edition (96 GB VRAM) with PyTorch 2.7.0+cu128 and batch size 8. We recorded 80 optimizer updates and discarded the first ten as warmup; CUDA timing was synchronized, and peak memory was measured with `torch.cuda.max_memory_allocated`.

Table 7 compares optimizer-update time and peak memory across five physical benchmarks. PMosFM has the lowest peak memory on four benchmarks and the second-lowest on turbulence reconstruction; its update time is lowest on Burgers and second-lowest on turbulence reconstruction. This pattern is consistent with separating feasibility from learned transport: feasible coordinates and the removal of terminal residual unrolling reduce memory and update cost relative to PBFM.

D.2 Benchmark summary

Table 6 summarizes the constraint forms covered by the experimental suite. The constrained-PDE reference suite combines affine initial or boundary constraints with linear or nonlinear conservation laws. The physics-constrained generation studies additionally cover compatible flow fields, algebraic closures, solver-embedded PDE decoders, and coupled constraints. A parameterization specifies feasible support, while each task separately declares the target law and physical evaluator.

Table 7: Training time (s per optimizer update) and peak allocated GPU memory (GB) across the benchmarks and baselines in Table 3. Lower is better; bold marks PMosFM.

Benchmark	Metric	FM-OT	CoCoGen	PIDM	DiffusionPDE	D-Flow	ECI	PCFM	PBFM	PMosFM
Dynamic Stall	Train s/iter ↓	0.0183	0.0090	0.0094	0.0511	0.0470	0.0169	0.0183	0.5440	**0.0563**
	Peak GB ↓	0.4027	0.9467	0.9465	2.1603	0.9475	0.4027	0.4027	34.1000	**0.2580**
Darcy Flow	Train s/iter ↓	0.0289	0.0073	0.0065	0.0517	0.0170	0.0276	0.0289	0.4520	**0.0185**
	Peak GB ↓	0.5621	0.2412	0.2412	2.1569	0.2418	0.5621	0.5621	114.0000	**0.1280**
Burgers	Train s/iter ↓	0.0246	0.0249	0.0245	0.0513	0.0240	0.0091	0.0091	0.0126	**0.0057**
	Peak GB ↓	2.2491	2.2465	2.2465	2.1565	2.2465	0.9487	0.9487	0.7245	**0.5676**
Kolmogorov Flow	Train s/iter ↓	0.0438	0.0088	0.0092	0.0940	0.0146	0.0420	0.0438	1.4200	**0.0208**
	Peak GB ↓	1.5317	0.9359	0.9359	2.1569	0.9350	1.5317	1.5317	33.2000	**0.1430**
Turbulence Flow Rec.	Train s/iter ↓	0.9988	2.0501	1.0805	0.0766	1.0249	1.0076	0.9988	5.9900	**0.9300**
	Peak GB ↓	9.7720	16.5738	16.5738	2.1596	16.5738	9.7720	9.7720	87.8000	**3.5000**

PBFM entries use the four-step Sum and PMosFM entries the one-step measurement.

Table 8: Benchmark summary. Physical checks assess conservation; statistical metrics assess coverage.

Benchmark	Task	Physical check	Distributional check
Darcy flow	Field generation	PDE/boundary residual	WD/JS; field moments
Dynamic stall	Conditional field	Wall-shear residual	MSE; shock statistics
Burgers	Nonlinear field	Conservation/boundary residual	WD/JS; shock location
Kolmogorov flow generation	Turbulence field	Incompressibility/flow statistics	MMD/WD; spectra; diversity
Kolmogorov flow reconstruction	Conditional inverse	Observation/flow consistency	MSE; CRPS/coverage
Turbulence forecast	Forecasting	Temporal/flow diagnostics	Forecast MSE; spectra; long-horizon stats
Turbulence reconstruction	Sparse inverse	Observation/flow consistency	MSE; CRPS/coverage; spectra

D.3 Zero-shot constrained-PDE reference

We compare sampling methods across PDEs with linear and nonlinear constraints under the common pretrained FFM backbone of the constrained-PDE protocol (Utkarsh et al., 2026). Table 8 pairs each task's physical checks with field or distributional metrics, separating constraint satisfaction from sample quality. Following Kerrigan et al. (2024) for pointwise functional statistics and Cheng et al. (2025) for constraint-aware feature-space evaluation, Table 9 reports pointwise mean MSE (MMSE), standard-deviation MSE (SMSE), Fréchet Poseidon distance (FPD), and constraint error (CE). MMSE and SMSE compare the generated and reference field moments, whereas FPD compares their feature distributions with a pretrained Poseidon encoder (Herde et al., 2024). For a constraint type $* \in \{\mathrm{IC}, \mathrm{BC}, \mathrm{CL}\}$, CE is the residual ℓ_2 norm over the relevant domain, averaged over N generated samples:

$$\mathrm{CE}(*) = \frac{1}{N} \sum_{n=1}^{N} \left\| \mathcal{R}_* \left(\hat{u}^{(n)} \right) \right\|_2 .$$

Lower values indicate closer agreement with the reference distribution or constraint. The shared baselines follow the left-to-right family order in Table 1, ending with PMosFM. Baselines FFM through PCFM report values from the established benchmark suite (Utkarsh et al., 2026). PBFM (Baldan et al., 2026) was retrained and evaluated under the identical benchmark protocol and data splits; its Navier–Stokes entry is unavailable because the upstream setup does not support the canonical 3D configuration. The final column reports PMosFM evaluated under the same independent test sets and protocol with one-step ($NFE = 1$) deterministic physical decoding.

Exact physical constraint preservation. Across five benchmark configurations and ten constraint evaluations, PMosFM reports numerical residuals below 10^{-5} for initial, boundary, and conservation constraints. For an exact decoder, the chart χ_c enforces the encoded discrete constraint within

Table 9: Zero-shot constrained-PDE benchmark. Heat and Navier–Stokes: linear IC/CL; Reaction–Diffusion and Burgers: nonlinear CL with IC/BC. Lower is better; best baseline in bold.

Dataset	Metric	FFM	Diffusion PDE	PDM	D-Flow	ECI	PCFM	PBFM	PMosFM
Heat Equation	MMSE / 10^{-2}	4.56	4.49	0.45	1.97	0.697	0.241	0.043	**0.016**
	SMSE / 10^{-2}	3.51	3.93	**0.02**	1.14	0.973	0.937	0.242	**0.003**
	CE (IC) / 10^{-2}	579	599	**0**	102	**0**	**0**	24.2	**0**
	CE (CL) / 10^{-2}	2.11	2.06	**0**	64.8	**0**	**0**	57.5	**0**
	FPD	1.77	1.70	0.17	2.70	1.34	1.22	–	**0.0806**
Navier–Stokes	MMSE / 10^{-2}	16.5	17.4	12.21	–	5.23	4.59	–	**4.41**
	SMSE / 10^{-2}	7.90	9.48	6.61	–	7.28	4.17	–	**3.24**
	CE (IC) / 10^{-2}	328	288	**0**	–	**0**	**0**	–	**0**
	CE (CL) / 10^{-2}	18.6	21.4	**0**	–	**0**	**0**	–	**0**
	FPD	2.81	3.70	2.01	–	1.04	**1.00**	–	1.01
Reaction–Diffusion IC	MMSE / 10^{-2}	2.92	3.16	1.74	0.318	0.324	0.026	**0.018**	0.042
	SMSE / 10^{-2}	2.54	2.54	0.43	6.86	0.060	0.583	0.017	**0.014**
	CE (IC) / 10^{-2}	445	451	**0**	215	**0**	**0**	33.4	**0**
	CE (CL) / 10^{-2}	3.87	3.82	**0**	29.7	6.00	**0**	25.3	**0**
	FPD	24.9	44.1	109	28.3	136	**15.7**	–	26.3
Burgers BC	MMSE / 10^{-2}	4.86	5.42	11.8	0.224	0.359	0.335	1.49	**0.120**
	SMSE / 10^{-2}	1.38	1.30	2.67	0.948	0.089	0.123	4.63	**0.0396**
	CE (BC) / 10^{-2}	409	426	**0**	95.7	20.3	**0**	102.3	**0**
	CE (CL) / 10^{-2}	6.91	6.20	**0**	15.0	15.7	**0**	35.5	**0**
	FPD	24.7	25.9	6.41	1.44	0.307	**0.292**	39.7	0.409
Burgers IC	MMSE / 10^{-2}	13.7	14.3	153	9.97	10.0	0.052	0.112	**0.030**
	SMSE / 10^{-2}	7.90	8.06	2.62	7.91	6.65	0.272	0.439	**0.019**
	CE (IC) / 10^{-2}	462	471	**0**	397	**0**	**0**	47.4	**0**
	CE (CL) / 10^{-2}	6.91	6.22	**0**	8.66	205	**0**	18.6	**0**
	FPD	33.5	35.8	99.7	22.1	1.31	**0.101**	2.20	0.283

D-Flow and PBFM Navier–Stokes and PBFM Heat/RD FPD are omitted (instabilities / no 1D protocol).

its valid domain. In contrast, physics-loss and penalty-based methods such as PBFM accumulate substantial residual violations along their trajectories; unconstrained flow and diffusion baselines drift above 1.

Distributional fidelity beyond feasibility. Across the evaluated equations, PMOSFM maintains strong agreement with the reference field statistics while preserving the encoded physical constraints. The consistent behavior of the moment-based metrics indicates that feasible-coordinate generation improves residual accuracy while allowing the learned transport to recover the dominant variability of the target distribution. Feature-space distances are more task dependent, so exact feasibility and distributional matching provide complementary measures of sample quality.

Task-dependent trade-offs. Reaction–Diffusion reveals complementary behavior across metrics: PMOSFM achieves improved constraint satisfaction and variance statistics alongside competitive mean-field accuracy. Similar variation in feature-space distances across tasks reinforces this metric dependence. Overall, PMOSFM preserves strict feasibility while remaining competitive across complementary measures of distributional fidelity.

D.3.1 BURGERS WITH FIXED INITIAL CONDITION

The Burgers IC rows of Table 9 are complemented by two visual diagnostics. Fig. 9 compares field moments with mass and boundary residuals, while Fig. 10 shows the evolving mean profiles and spread. Both PCFM and PMosFM have small mass and boundary residuals in Fig. 9, but their field statistics differ. PMosFM more closely follows the reference mean profiles and standard-deviation pattern in Fig. 10.

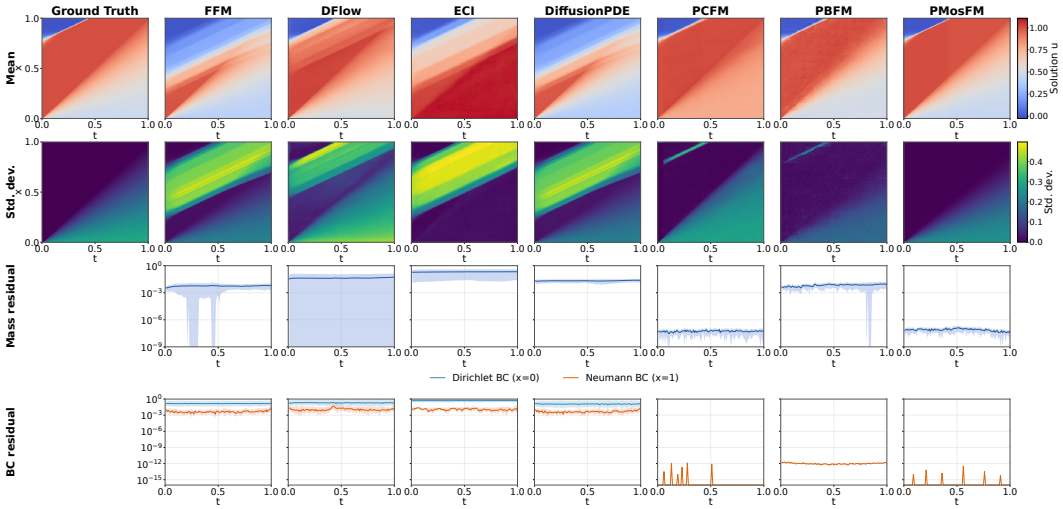


Figure 9: Fixed-IC Burgers conservation diagnostics. Rows compare solution means and standard deviations, followed by mass and boundary residuals for the displayed methods.

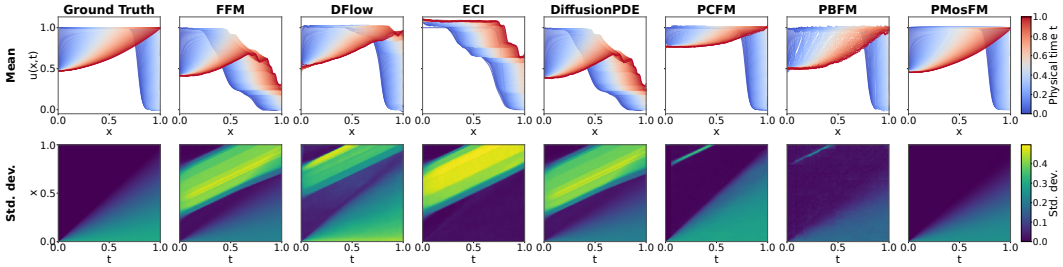


Figure 10: Fixed-IC Burgers solution statistics. Mean solution profiles across time and pointwise standard deviations for the reference and displayed methods.

D.4 Feasible-set and target-law study

We evaluate the closed-form ellipse construction of Section 4.5 over five independent seeds and 24,000 samples per seed. In Table 10, both exact parameterizations give residual RMS near 10^{-16}, but the co-area law lowers KL from 0.369 to 0.00207 relative to the volume law. Thus, exact feasibility alone does not recover the target law.

Table 10: Residual-manifold support and measure over five independent seeds (24,000 samples each). Exact parameterizations yield near-zero residual; co-area correction recovers the target measure.

Method	Residual RMS ↓	$\mathrm{KL}(q\|q^\star)$ ↓	$\mathrm{TV}(q, q^\star)$ ↓
Ambient soft residual	$7.120\times10^{-2} \pm 2.264\times10^{-4}$	$2.077\times10^{-3} \pm 1.548\times10^{-4}$	$2.358\times10^{-2} \pm 8.234\times10^{-4}$
Exact parameterization, volume law	$1.917\times10^{-16} \pm 1.326\times10^{-18}$	$3.689\times10^{-1} \pm 2.069\times10^{-3}$	$3.723\times10^{-1} \pm 1.229\times10^{-3}$
Exact parameterization, co-area law	$1.195\times10^{-16} \pm 1.646\times10^{-18}$	$2.071\times10^{-3} \pm 1.448\times10^{-4}$	$2.468\times10^{-2} \pm 1.634\times10^{-3}$

Construction and representation scope. Exact support fixes the feasible set, whereas the constrained measure specifies the probability law on that set. The controlled feasible set is the ellipse $\chi(\theta) = (1.65\cos\theta, 0.72\sin\theta)$ with scalar residual

$$R(x) = \exp(1.2\cos\theta(x))\left(\frac{x_1^2}{1.65^2} + \frac{x_2^2}{0.72^2} - 1\right),$$

where $\theta(x)$ is the ellipse angle. The residual scaling preserves the zero set while making $\|J_R\|$ nonuniform. For a tubular ambient density constant on the ellipse, the co-area target is $q^\star(\theta) \propto$

$\|\partial_\theta \chi\| \,/\, \|J_R\|$, while a volume-only exact sampler has $q_{\mathrm{vol}}(\theta) \propto \|\partial_\theta \chi\|$. For an encoded exact residual, the controlled objective in Eq. 13 retains the physical matching terms after residual factorization removes the corresponding residual-normal correction. The held-out physical studies separately examine encode–decode error, encoded residual, and independently evaluated residual.

D.5 Held-out diagnostics

This held-out pre-training study reports encode–decode field error together with decoder-side and independently evaluated residuals. The evaluation is performed with an evaluator that is independent of the coordinate decoder. The manifold-conditioned Dynamic Stall case closes the specified algebraic constraint under both evaluators. For Kolmogorov flow, the central-difference implementation has external RMS 1.353×10^{-1}, whereas the spectral implementation has external RMS 2.422×10^{-6}, with both implementations retaining relative ℓ_2 error near 0.92.

Table 11: Population tangent-regression conditioning with 64 eigenmodes and the same stable step rule. Lower is better.

Coordinates	Active κ_+	Steps to $90\% \downarrow$
Raw covariance coordinates	10^6	2,558,427
Exact state-whitened coordinates	1	1

D.6 Tangent-regression conditioning

Table 18 and the left panel of Fig. 17 show that grid refinement worsens raw conditioning, while the ideal residual-space metric keeps the active condition number near one. The population study uses 64 eigenmodes with a fixed step rule. Table 11 and the right panel of Fig. 17 show that exact whitening removes covariance anisotropy and accelerates convergence in the stated population-regression setting.

D.7 Dynamic Stall field and ensemble results

Fig. 11 compares the methods for one held-out Dynamic Stall condition. The PMosFM row places the shock-associated transition at the same location across the pressure, velocity-gradient, temperature, density, and wall-shear channels as the reference field.

Fig. 12 compares the conditional ensemble means and standard deviations of the six generated channels with the reference statistics. The PMosFM mean preserves the dominant shock-associated structure, but its standard-deviation maps show weaker fluctuations than the reference under the shared color scales.

D.8 Held-out one-step turbulence

The Q–R evaluator uses periodic Fourier differentiation followed by trace removal. Let A denote the resulting velocity-gradient tensor, with $Q = -\frac{1}{2}\,\mathrm{tr}(A^2)$ and $R = -\det(A)$. Fixed histogram bins and the same normalization are used for all methods. Table 12 shows that PMosFM has the lowest field MSE, divergence RMS, and all three Q–R distribution distances among the evaluated methods.

Figs. 13 and 14 compare generated and held-out Q–R laws under the common evaluator. PMosFM follows the reference contours more closely than PCFM, ECI, and PBFM, while FM-OT, FluidFlow, and PCFT also retain much of the contour structure in the full-baseline comparison.

D.9 Qualitative turbulence diagnostics

Fig. 15 shows that reconstruction becomes harder as observations become sparser. At 99% missingness, PMosFM retains a coarse spatial pattern, but its spectra and gradient PDFs still differ from the reference.

The held-out HIT benchmark evaluates one-step generation, while the random-mask test assesses sparse reconstruction. PMosFM attains the lowest field MSE, divergence RMS, and Q–R distribution distances in the HIT table; in reconstruction, it has the lowest missing-region RMSE at all three rates and retains an identifiable large-scale pattern at 99% missingness (Tables 12 and 13; Fig. 15).

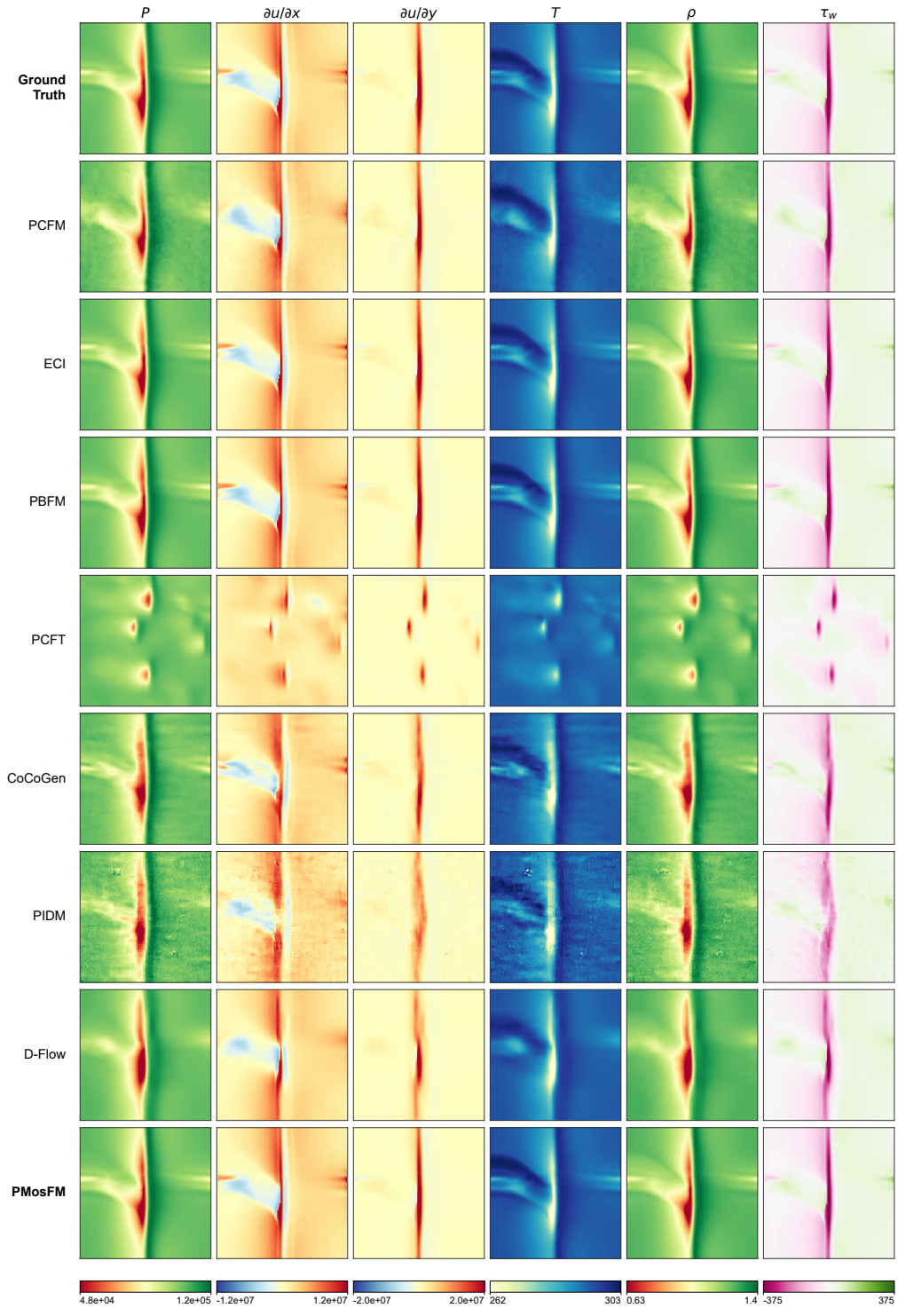


Figure 11: Qualitative Dynamic Stall comparison. One held-out condition is shown. Columns show pressure, velocity gradients, temperature, density, and wall shear. Rows compare the reference and model predictions; PMosFM retains the shock-transition location and cross-channel structure.

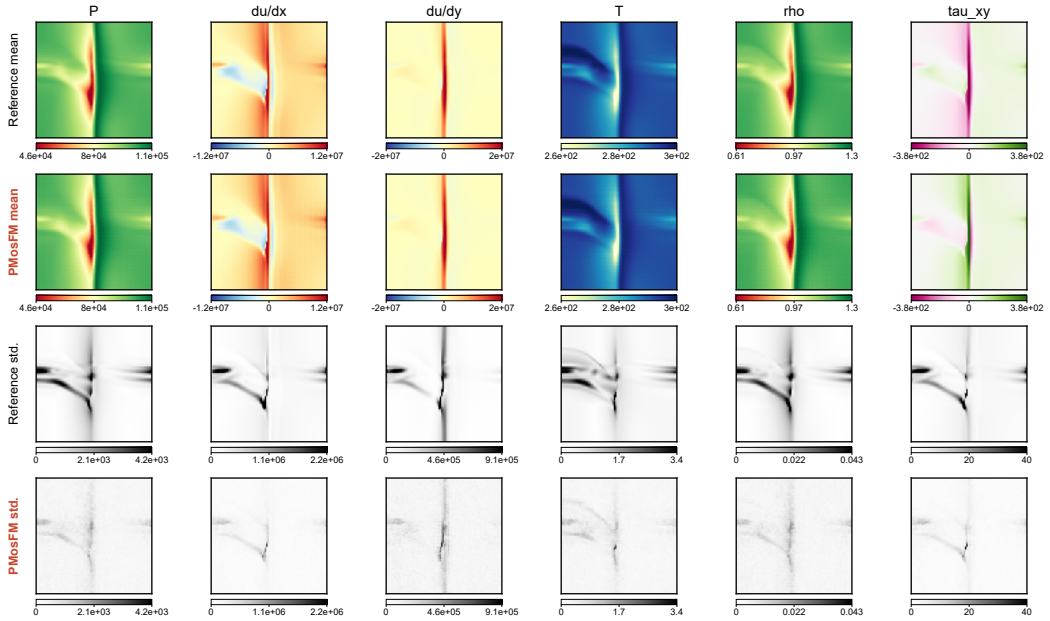


Figure 12: Dynamic Stall ensemble moments. Conditional means and standard deviations of pressure, tangential velocity gradients, temperature, density, and wall shear for one held-out condition.

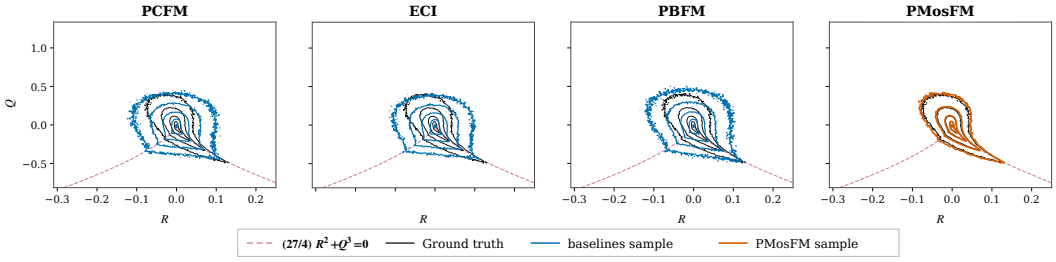


Figure 13: Focused Q–R comparison for turbulence forecasting. Selected physics-constrained baselines are compared with the held-out ground truth.

E Additional Ablations

Loss balance. Table 14 tests whether velocity supervision and endpoint matching play complementary roles. On both Kolmogorov and Darcy flow, the balanced weight gives the lowest field error and physical sliced W_2, together with a near-zero reported residual. Moving toward either single-term objective worsens all three metrics, suggesting that the two losses are complementary.

Table 14: Velocity–endpoint loss balance on Kolmogorov and Darcy flow.

Benchmark	Configuration	Loss Formulation	Field MSE ↓	Residual / Div. RMS ↓	Physical Sliced W_2 ↓
Kolmogorov Flow (128×128)	PE only ($\gamma = 0.0$)	Direct Endpoint $\mathcal{L}_{\mathrm{PE}}$	0.4241 ± 0.0182	1629.9 ± 52.4	0.3852 ± 0.0124
	PE-heavy ($\gamma = 0.25$)	Two-Time Joint	0.1853 ± 0.0118	850.4 ± 38.1	0.2821 ± 0.0092
	PMosFM Balanced ($\gamma = 0.50$)	**Two-Time Joint** ($\mathcal{L}_{\mathrm{FM}} + \mathcal{L}_{\mathrm{PE}}$)	$\mathbf{0.0531 \pm 0.0019}$	$\mathbf{3.45 \times 10^{-9}}$	$\mathbf{0.2522 \pm 0.0014}$
	FM-heavy ($\gamma = 0.75$)	Two-Time Joint	0.2184 ± 0.0145	1453.1 ± 46.2	0.2789 ± 0.0101
	FM only ($\gamma = 1.0$)	Velocity Flow Matching $\mathcal{L}_{\mathrm{FM}}$	0.4714 ± 0.0210	7258.3 ± 112.5	0.3621 ± 0.0153
Darcy Flow (64×64)	PE only ($\gamma = 0.0$)	Direct Endpoint $\mathcal{L}_{\mathrm{PE}}$	0.0382 ± 0.0024	870.8 ± 42.1	0.0284 ± 0.0018
	PE-heavy ($\gamma = 0.25$)	Two-Time Joint	0.0115 ± 0.0008	128.4 ± 9.5	0.0098 ± 0.0006
	PMosFM Balanced ($\gamma = 0.50$)	**Two-Time Joint** ($\mathcal{L}_{\mathrm{FM}} + \mathcal{L}_{\mathrm{PE}}$)	$\mathbf{0.0044 \pm 0.0001}$	$\mathbf{(1.41 \pm 0.03) \times 10^{-12}}$	$\mathbf{0.0044 \pm 0.0001}$
	FM-heavy ($\gamma = 0.75$)	Two-Time Joint	0.0092 ± 0.0007	240.1 ± 14.2	0.0087 ± 0.0005
	FM only ($\gamma = 1.0$)	Velocity Flow Matching $\mathcal{L}_{\mathrm{FM}}$	0.0415 ± 0.0031	1577.2 ± 68.3	0.0312 ± 0.0021

Note. Results are mean $\pm$ standard deviation over five independent seeds.

Table 12: Held-out turbulence evaluation for physics-constrained generation uses a common evaluator with fixed Q–R bins. All methods follow the main-text objective; lower is better.

Method	Field MSE ↓	Divergence RMS ↓	Q–R JS ↓	Q–R TV ↓	Hellinger ↓
FM-OT	1.097	1.000×10^{-1}	9.165×10^{-3}	7.499×10^{-2}	1.048×10^{-1}
advNO	2.205	1.248	1.317×10^{-1}	4.171×10^{-1}	3.798×10^{-1}
FluidFlow	1.102	1.004×10^{-1}	9.422×10^{-3}	7.609×10^{-2}	1.062×10^{-1}
CoCoGen	2.006×10^{2}	4.693	6.803×10^{-2}	2.857×10^{-1}	2.727×10^{-1}
DiffusionPDE	1.785	6.874×10^{-1}	1.815×10^{-1}	4.653×10^{-1}	4.617×10^{-1}
PIDM	1.695×10^{2}	4.248	6.351×10^{-2}	2.721×10^{-1}	2.638×10^{-1}
D-Flow	2.048×10^{2}	4.676	6.107×10^{-2}	2.664×10^{-1}	2.588×10^{-1}
ECI	6.472×10^{-2}	1.426×10^{-1}	8.684×10^{-2}	3.360×10^{-1}	3.041×10^{-1}
PCFM	6.357×10^{-2}	7.464×10^{-2}	8.099×10^{-2}	3.227×10^{-1}	2.936×10^{-1}
PCFT	1.096	1.004×10^{-1}	9.230×10^{-3}	7.338×10^{-2}	1.053×10^{-1}
PBFM	4.892×10^{-2}	3.387×10^{-2}	5.793×10^{-2}	2.708×10^{-1}	2.478×10^{-1}
PMosFM	$\mathbf{5.359 \times 10^{-3}}$	$\mathbf{1.527 \times 10^{-7}}$	$\mathbf{5.961 \times 10^{-3}}$	$\mathbf{5.293 \times 10^{-2}}$	$\mathbf{8.513 \times 10^{-2}}$

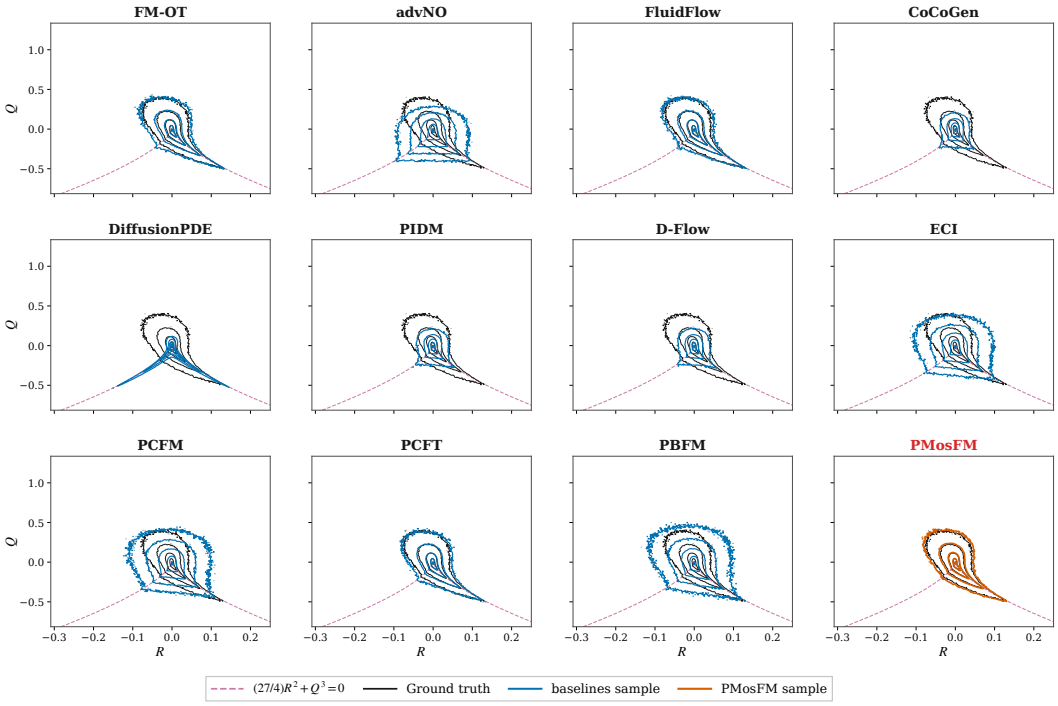


Figure 14: Full Q–R comparison for turbulence forecasting. All evaluated methods are compared with the held-out ground truth.

Endpoint-gradient term. The held-out turbulence implementation is averaged over five independently trained model states and five independent inference seeds, whereas external methods are single-state estimates. The held-out turbulence implementation has 948,868 parameters and performs one neural evaluation followed by physical decoding. It trains with direct velocity MSE, decoded-endpoint MSE, and a 0.2-weighted endpoint-gradient MSE under the hard projector. Table 15 isolates the endpoint-gradient term in this one-step implementation. Velocity and Endpoint denote direct velocity MSE and decoded-endpoint MSE, respectively; λ_{∇} is the endpoint-gradient MSE coefficient, and "–" denotes omission. Both variants share a residual-consistent projector and are evaluated over 16 validation windows, five independent training seeds, and five independent inference seeds. Bold marks the lower value in each diagnostic column. Table 15 shows a small reduction in physical MSE (5.631×10^{-3} to 5.524×10^{-3}) and lower Q–R distributional distances, while divergence RMS remains near 1.5×10^{-7}.

Table 13: TFR sparse-reconstruction metrics for held-out frames 168–186 (32 samples per condition). Lower is better except interval width; bold marks the minimum at each missing rate.

Missing	Method	Missing-region RMSE ↓	Physical residual ↓	CRPS ↓	Coverage error ↓	Interval width
50%	PBFM	0.063159	1.102126	0.024882	0.075232	0.084228
	ECI	0.045158	1.910683	0.019637	0.035143	0.126545
	PCFM	0.067615	1.520453	0.030269	0.099867	0.112164
	PCFT	0.069091	1.081061	0.027396	0.095131	0.082924
	PMosFM	**0.029653**	**0.620392**	**0.008742**	**0.010382**	0.036809
90%	PBFM	0.151450	0.677789	0.075305	0.371066	0.090451
	ECI	0.066441	1.867086	0.028625	0.004689	0.158808
	PCFM	0.089082	1.211505	0.038495	0.070901	0.148736
	PCFT	0.152720	0.698189	0.075118	0.351275	0.088116
	PMosFM	**0.045076**	**0.628506**	**0.014650**	**0.036005**	**0.059326**
99%	PBFM	0.187398	1.594082	0.095594	0.430716	0.092738
	ECI	0.165424	1.695730	0.078446	0.173864	0.192676
	PCFM	0.177361	1.103668	0.085315	0.219686	0.171779
	PCFT	0.187990	0.622136	0.095010	0.416529	0.090029
	PMosFM	**0.075462**	**0.53680**	**0.026655**	**0.021873**	0.119706

Table 15: One-step objective ablation for one-step turbulence forecasting. Lower is better.

	Training terms			Held-out diagnostic				
Variant	Velocity	Endpoint	λ_∇	Physical MSE ↓	Divergence RMS ↓	Q–R JS ↓	Q–R TV ↓	Hellinger ↓
No gradient	yes	yes	–	5.631×10^{-3}	1.507×10^{-7}	6.257×10^{-3}	5.634×10^{-2}	8.698×10^{-2}
Full objective	yes	yes	0.2	5.524×10^{-3}	1.508×10^{-7}	6.202×10^{-3}	5.565×10^{-2}	8.664×10^{-2}

Preconditioning. Table 16 compares identity, diagonal standardization, and preconditioning. In the controlled regression, preconditioning gives the lowest condition number and iteration count; it also gives the lowest reported Kolmogorov divergence error.

Table 16: Geometric preconditioning effects on conditioning and Kolmogorov-flow convergence.

Transform	**Condition $\kappa(H)$ ↓**	**90% Reduction ↓**	**Kolmogorov Div Err ↓**	**Iteration ratio**
Identity (Vanilla)	1.00×10^6	2,558,427	625.17	1.0× (baseline)
Standardization (Diagonal)	4.82×10^3	1,240	485.30	2,063×
Preconditioning (PMosFM)	**1.00**	**1** step	**234.01**	$\mathbf{> 10^6}\times$

Decoder iterations. Table 17 examines the Darcy decoder's accuracy–cost trade-off. Higher K lowers both errors but raises decoding latency, which is separate from one network evaluation.

Table 17: Darcy decoder accuracy and latency across iteration counts.

Iterations K	**Relative Pressure Error ↓**	**Independent Residual RE ↓**	**Latency / sample (ms) ↓**
16	0.0124 ± 0.0012	0.0385 ± 0.0035	**0.62 ms**
32	0.0031 ± 0.0004	0.0084 ± 0.0009	0.98 ms
64	0.0008 ± 0.0001	0.0019 ± 0.0002	1.65 ms
128	0.0002 ± 0.00003	0.00048 ± 0.00005	2.84 ms
256	$\mathbf{0.00004 \pm 0.00001}$	$\mathbf{0.00012 \pm 0.00002}$	5.12 ms

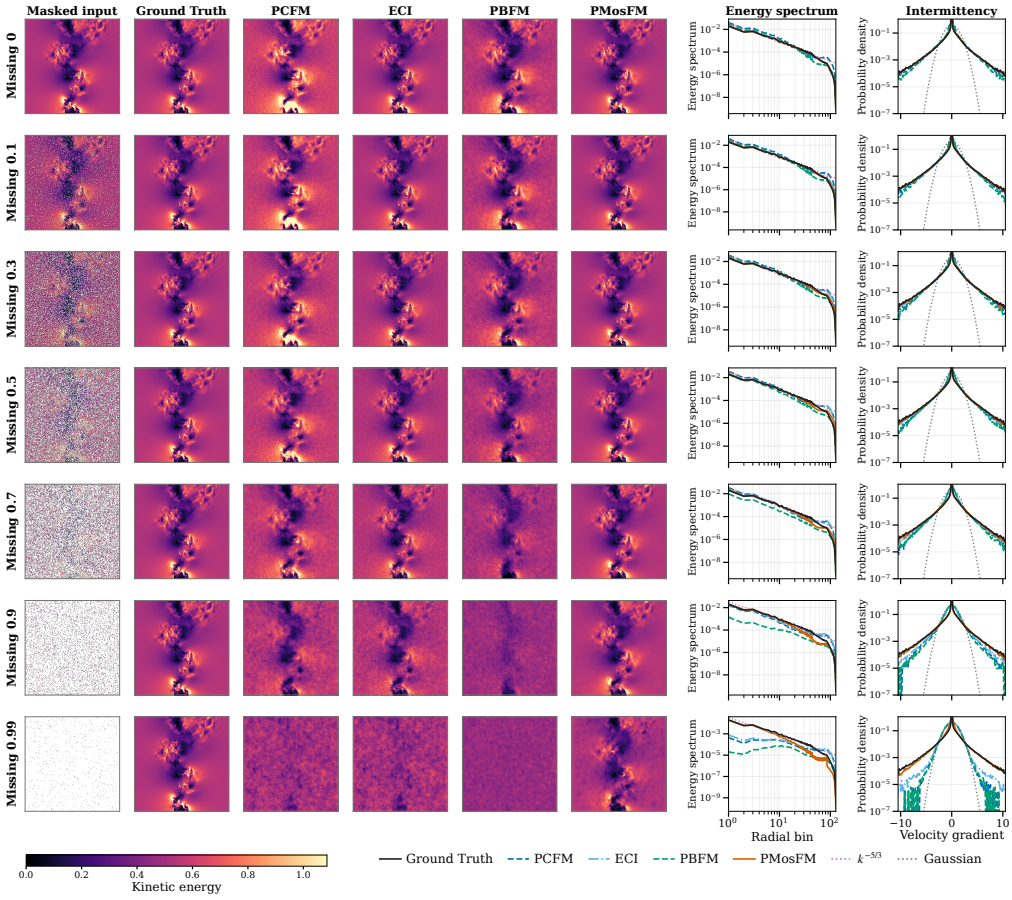


Figure 15: Random-mask reconstruction. One held-out turbulence frame is tested with shared exact-count velocity masks (0–99% missing). Columns show the input, ground truth, PCFM, ECI, PBFM, and PMosFM; lower panels show energy spectra and standardized $\partial_z w$ PDFs.

F Mathematical Setting and Assumptions

This appendix records assumptions, proofs, and constructions for the discretized residual R_h, proceeding from residual factorization through preconditioning to two-time matching.

Fix a condition c and suppress the condition index when no ambiguity arises. Let $a \in \mathbb{R}^s$ denote either a generator input, a latent variable, or a vector of model parameters. For a positive semidefinite matrix M, write $\|v\|_M^2 = v^\top M v$. For a matrix A, let $\sigma_{\min}^+(A)$ denote the smallest nonzero singular value of A and define the active-subspace condition number

$$\kappa_+(A) := \frac{\sigma_{\max}(A)}{\sigma_{\min}^+(A)}.$$

This convention separates spectral stiffness from null directions imposed by a physical constraint.

Assumption 1 (Regular residual factorization)**.** On the represented target region, $R_h(\cdot, c)$ is C^1 and $D_x R_h(x, c)$ has constant rank q. The map $\chi_c : \mathcal{Y}_c \to \mathcal{X}_h$ is a C^1 local embedding of dimension $m = d - q$ on each active coordinate patch; the union of represented feasible sets covers the target region, and

$$R_h(\chi_c(y), c) = 0 \qquad \text{for every } y \in \mathcal{Y}_c. \tag{17}$$

The represented component is $\mathcal{M}_c^\chi = \chi_c(\mathcal{Y}_c) \subseteq \mathcal{Z}_c$ from Eq. 1 on each patch. Every intrinsic model path remains inside a valid coordinate domain or is covered by compatible local parameterizations.

Assumption 2 (Differentiability and finite moments)**.** The two-time map $T_\theta^{s,t}$ and the decoder Ψ_c are differentiable in the variables under analysis. The interpolation covariance defining $P_{s,c}$ is fi-

nite, symmetric, and positive semidefinite. Whenever a second-order expansion is used, each map appearing in the expansion is C^2 on the corresponding line segment.

F.1 Baseline constraint mechanisms

Relative to the ambient objective in Eq. 3, PBFM defines separate flow-matching and physical losses, denoted $\mathcal{J}_{\mathrm{FM}}$ and $\mathcal{J}_R$, and coordinates their gradients before terminal residual evaluation (Baldan et al., 2026):

$$g_{\mathrm{FM}} = \nabla_\phi \mathcal{J}_{\mathrm{FM}}, \qquad g_R = \nabla_\phi \mathcal{J}_R, \qquad g_{\mathrm{update}} = \mathrm{ConFIG}(g_{\mathrm{FM}}, g_R).$$

PCFT instead applies weak-form fine-tuning (Tauberschmidt et al., 2026). DiffusionPDE adds sampling-time physics guidance (Huang et al., 2024); CCFM applies a chance-constrained projection (Liang et al., 2025). ECI extrapolates a clean endpoint and applies a correction $\mathcal{Q}_c$ before mapping back to the current noise level (Cheng et al., 2025):

$$\widehat{x}_{1|\tau} = x_\tau + (1-\tau) v_\phi(x_\tau, \tau, c), \qquad x^c_{1|\tau} = \mathcal{Q}_c(\widehat{x}_{1|\tau}).$$

PCFM applies nonlinear corrections to estimated terminal states during sampling (Utkarsh et al., 2026). For a full-row-rank residual Jacobian J_R, a local Gauss–Newton correction contains

$$x_1^{\mathrm{GN}} = \widetilde{x}_1 - J_R^\top (J_R J_R^\top)^{-1} R_h(\widetilde{x}_1, c),$$

followed by method-specific back-mapping, relaxed updates, or a constrained solve when required.

G One-Step Sensitivity and Curvature

G.1 Three notions of a step

Training-time unrolling (Baldan et al., 2026) evaluates a nonlinear residual after intermediate updates; inference-time integration counts network evaluations used to solve a learned ODE. PMosFM learns a finite-interval map evaluated once at inference. The Euler rollout in Eq. 18 serves as a conditioning diagnostic.

For an n-step Euler rollout from sampling time t, define

$$\begin{aligned} \mathcal{U}_i(x) &= x + \Delta\tau\, v_\theta(x, \tau_i, c), \quad \Delta\tau = \frac{1-t}{n}, \quad \tau_i = t + i\Delta\tau, \\ T^{(n)}_{\theta, t\to 1} &= \mathcal{U}_{n-1} \circ \cdots \circ \mathcal{U}_0. \end{aligned} \tag{18}$$

Let $A_i = D_x v_\theta(x^{(i)}, \tau_i, c)$. The chain rule gives

$$D_x T^{(n)}_{\theta, t\to 1} = (I + \Delta\tau A_{n-1}) \cdots (I + \Delta\tau A_0). \tag{19}$$

Proposition 2 (Finite-interval sensitivity)**.** *Let $x(s)$ solve $\dot{x}(s) = v(x(s), s, c)$ for $s \in [t, 1]$, and assume $A(s) = D_x v(x(s), s, c)$ is continuous. The endpoint sensitivity $J_{t,s} = D_{x(t)} x(s)$ satisfies*

$$\frac{dJ_{t,s}}{ds} = A(s) J_{t,s}, \qquad J_{t,t} = I. \tag{20}$$

The sensitivity also obeys

$$\|J_{t,1}\|_2 \le \exp\left(\int_t^1 \|A(s)\|_2 \, ds \right). \tag{21}$$

If the flow is locally invertible, then

$$\kappa_2(J_{t,1}) \le \exp\left(2 \int_t^1 \|A(s)\|_2 \, ds \right). \tag{22}$$

Proof. Differentiating with respect to the initial state gives Eq. 20; Gronwall's inequality yields Eq. 21. Eq. 19 discretizes Proposition 2; the controlled study evaluates the conditioning of the one-step endpoint Jacobian. The inverse sensitivity satisfies the corresponding backward variational equation, and the two norm bounds give Eq. 22. □

G.2 Residual-normal curvature

Let $G_\phi(a, c)$ be any differentiable endpoint generator and define

$$r_\phi(a, c) = R_h(G_\phi(a, c), c), \qquad \ell_R(a) = \frac{1}{2} \|r_\phi(a, c)\|^2_{W_R}, \quad W_R \succeq 0. \tag{23}$$

The residual R_h has n_R components.

Lemma 3 (Exact Hessian decomposition)**.** *Assume R_h and G_ϕ are C^2. Let $J_G = D_a G_\phi$ and $J_R = D_x R_h|_{G_\phi(a,c)}$. Then*

$$\begin{aligned} \nabla_a \ell_R &= (J_R J_G)^\top W_R r_\phi, \\ \nabla^2_a \ell_R &= (J_R J_G)^\top W_R (J_R J_G) + \sum_{j=1}^{n_R} (W_R r_\phi)_j \, \nabla^2_a (r_\phi)_j. \end{aligned} \tag{24}$$

At a feasible output, $r_\phi = 0$, the second term vanishes and the exact Hessian equals the Gauss–Newton block

$$H^R_{\mathrm{GN},a} = J_G^\top J_R^\top W_R J_R J_G. \tag{25}$$

Proof. Apply the chain rule to Eq. 23. Differentiating the gradient gives the two terms in Eq. 24. Feasibility eliminates the latter. □

Corollary 4 (Squared active spectral spread)**.** *Let $B = W_R^{1/2} J_R J_G$. On the active subspace of B,*

$$H^R_{\mathrm{GN},a} = B^\top B, \qquad \kappa_+(H^R_{\mathrm{GN},a}) = \kappa_+(B)^2. \tag{26}$$

In addition,

$$\lambda_{\max}(H^R_{\mathrm{GN},a}) \le \|W_R\|_2 \, \|J_R\|_2^2 \, \|J_G\|_2^2 .$$

Proof. Eq. 26 follows by squaring the singular values of B; submultiplicativity bounds the norm. □

Remark 5 (Condition-number product). The common heuristic $\kappa(H_{\mathrm{GN}}) \approx \kappa(J_G)^2 \kappa(J_R^\top J_R)$ depends on alignment between the relevant singular subspaces. Corollary 4 gives the unconditional result: the condition number of the composed residual differential is squared. After generator composition, exact residual factorization also removes the residual block in Eq. 25.

H Manifold Parameterization

H.1 Residual-factorization differential

Lemma 6 (Residual-factorization differential)**.** *Under Assumption 1, for every $y \in \mathcal{Y}_c$ and $x = \chi_c(y)$,*

$$\begin{aligned} J_R(x, c) J_\chi(y, c) = 0, \qquad & \operatorname{range} J_\chi(y, c) \subseteq \ker J_R(x, c), \\ \operatorname{range} J_\chi(y, c) = \ker J_R(x, c) = T_x \mathcal{M}^\chi_c. & \end{aligned} \tag{27}$$

Proof. Differentiating $R_h(\chi_c(y), c) = 0$ gives $J_R J_\chi = 0$, which implies $\operatorname{range} J_\chi \subseteq \ker J_R$. Both spaces in Eq. 27 have dimension $d - q$, so the inclusion is equality. □

H.2 Exact endpoint feasibility

Let $T^{0,1}_\phi(r_0, c)$ lie in the valid coordinate domain and set $G^{\mathrm{RF}}_{\phi,c} = \Psi_c \circ T^{0,1}_\phi$. Assumption 1 gives

$$R_h(G^{\mathrm{RF}}_{\phi,c}(r_0, c), c) = R_h\Big(\chi_c\Big(\mu_c + C_c^{-1} T^{0,1}_\phi(r_0, c)\Big), c\Big) = 0. \tag{28}$$

The left-hand side of Eq. 28 is identically zero as a function of r_0 and of every model parameter entering $T^{0,1}_\phi$. Differentiation with respect to any such variable a yields

$$D_a[R_h \circ G^{\mathrm{RF}}_{\phi,c}] = J_R J_\chi C_c^{-1} D_a T^{0,1}_\phi = 0. \tag{29}$$

Eq. 29 shows that the encoded residual contributes no Gauss–Newton block on the valid coordinate domain. This local identity does not control the full endpoint Jacobian or global optimization.

Corollary 7 (Feasibility of a decoded intrinsic path)**.** *Let $y : [0,1] \to \mathcal{Y}_c$ be any path and define $x(s) = \chi_c(y(s))$. Under Assumption 1,*

$$R_h(x(s), c) = 0 \qquad \text{for every } s \in [0,1]. \tag{30}$$

If y is differentiable, then $\dot{x}(s) = J_\chi(y(s), c)\dot{y}(s) \in T_{x(s)}\mathcal{M}_c^\chi$.

Proof. Eq. 17 gives the feasibility in Eq. 30; differentiating the path and applying Lemma 6 gives tangency. □

H.3 Intrinsic-coordinate penalty block

For an ambient linear residual $R_h(x) = A_h x - b$, the penalized objective and the Hessian are

$$\mathcal{L}_\lambda(x) = \mathcal{L}_{\text{gen}}(x) + \frac{\lambda}{2}\|A_h x - b\|_W^2, \qquad \nabla^2 \mathcal{L}_\lambda = H_{\text{gen}} + \lambda A_h^\top W A_h. \tag{31}$$

Theorem 8 (Penalty-block scaling under exact intrinsic coordinates)**.** *Write $H_{\lambda,h} = H_{\text{gen}} + \lambda A_h^\top W A_h$. Let $x = x_p + N_h y$ with $A_h x_p = b$, $A_h N_h = 0$, and $N_h^\top N_h = I$. Then the intrinsic Hessian is $N_h^\top H_{\text{gen}} N_h$ and is independent of λ. If A_h discretizes an order-p operator with $\sigma_{\max}(A_h) = \Theta(h^{-p})$, $\sigma_{\min}^+(A_h) = \Theta(1)$, and W is uniformly spectrally equivalent to the identity on the active residual space, the largest eigenvalue contributed by the residual penalty block is $\Theta(\lambda h^{-2p})$. For an ambient condition-number statement, fix a declared subspace V_h and interpret restriction with an orthonormal basis of V_h. Suppose $H_{\text{gen}} \succeq 0$, the restriction $H_{\lambda,h}|_{V_h}$ is positive definite, $\sigma_{\max}(A_h|_{V_h}) = \Theta(h^{-p})$, and $V_h \cap \ker A_h$ contains a unit vector z_h with $0 < z_h^\top H_{\text{gen}} z_h \le L_{\tan}$ for an h- and λ-independent constant $L_{\tan}$. Then*

$$\kappa(H_{\lambda,h}|_{V_h}) \ge \frac{c_W \lambda \sigma_{\max}(A_h|_{V_h})^2}{L_{\tan}} = \Omega(\lambda h^{-2p}), \tag{32}$$

where $c_W I \preceq W$ on the active residual space. Writing κ_+ for the ratio of the largest to the smallest positive eigenvalue, if a tangent preconditioner S satisfies

$$(1-\delta)I \preceq S^\top N_h^\top H_{\text{gen}} N_h S \preceq (1+\delta)I, \qquad 0 \le \delta < 1, \tag{33}$$

then

$$\kappa_+\big(S^\top N_h^\top H_{\text{gen}} N_h S\big) \le \frac{1+\delta}{1-\delta}, \tag{34}$$

independently of λ.

Substituting $x = x_p + N_h y$ into Eq. 31 gives $A_h x - b = A_h N_h y = 0$. Differentiating the restricted generative objective twice gives the intrinsic Hessian $N_h^\top H_{\text{gen}} N_h$; the residual term contributes $\lambda N_h^\top A_h^\top W A_h N_h = 0$ for every λ.

Uniform spectral equivalence of W implies $0 < c_W \le C_W < \infty$, independent of h, such that

$$c_W \|A_h z\|_2^2 \le z^\top A_h^\top W A_h z \le C_W \|A_h z\|_2^2.$$

On the active normal subspace, the assumed singular-value scaling gives

$$\lambda_{\max}(\lambda A_h^\top W A_h) = \Theta(\lambda h^{-2p}), \qquad \lambda_{\min}^+(\lambda A_h^\top W A_h) = \Theta(\lambda).$$

This establishes the claimed normal stiffness scaling. Under the additional hypotheses of Theorem 8, positive semidefiniteness gives

$$\lambda_{\max}(H_{\lambda,h}|_{V_h}) \ge c_W \lambda \sigma_{\max}(A_h|_{V_h})^2.$$

For the unit vector $z_h \in V_h \cap \ker A_h$ in the theorem,

$$\lambda_{\min}(H_{\lambda,h}|_{V_h}) \le z_h^\top H_{\lambda,h} z_h = z_h^\top H_{\text{gen}} z_h \le L_{\tan}.$$

Dividing proves Eq. 32. The ambient bound applies on the declared subspace and does not remove generator-dependent tangential curvature. Finally, the Loewner bounds in Eq. 33 place every active eigenvalue of the preconditioned intrinsic Hessian in $[1-\delta, 1+\delta]$; taking their ratio proves Eq. 34. The matrix S is an abstract reference transform satisfying Eq. 33; the implemented transforms C_c and $P_{s,c}$ are analyzed separately in Proposition 13.

For the ideal residual-space Riesz choice $W_h = (A_h A_h^\top + \varepsilon I)^{-1}$, the nonzero eigenvalues of $A_h^\top W_h A_h$ are $\sigma_i^2/(\sigma_i^2 + \varepsilon)$. They approach one when ε is small relative to the active spectrum. For small ε relative to the active spectrum, this balances the residual-normal block, while intrinsic parameterization and tangent conditioning control the remaining feasible directions.

H.4 Approximate parameterizations and error bounds

For an approximate decoder, define two independent approximation errors

$$\begin{aligned}\varepsilon_{\text{chart}} &\coloneqq \sup_{y\in\mathcal{Y}_c} \|R_h(\chi_c(y),c)\|_2\,,\\ \delta_{\text{chart}} &\coloneqq \sup_{y\in\mathcal{Y}_c} \|D_y[R_h(\chi_c(y),c)]\|_2\,.\end{aligned} \tag{35}$$

where $\varepsilon_{\text{chart}}$ controls residual value, whereas δ_{chart} controls the pulled-back residual differential.

Theorem 9 (Approximate feasibility and residual Gauss–Newton-block bound)**.** *Suppose the quantities in Eq. 35 are finite and $T_\phi^{0,1}(r_0,c)$ maps every evaluated source coordinate into the coordinate domain. For $G_{\phi,c}^{\mathrm{RF}}=\Psi_c\circ T_\phi^{0,1}$,*

$$\left\|R_h(G_{\phi,c}^{\mathrm{RF}}(r_0,c),c)\right\|_2 \le \varepsilon_{\text{chart}}. \tag{36}$$

For any differentiable variable a and $W_R\succeq 0$,

$$\left\|H_{\mathrm{GN},a}^{R}\right\|_2 \le \|W_R\|_2\ \delta_{\text{chart}}^2\left\|C_c^{-1}D_aT_\phi^{0,1}\right\|_2^2. \tag{37}$$

Proof. Eq. 36 follows from the first supremum in Eq. 35. The chain rule gives

$$D_a[R_h\circ\Psi_c\circ T_\phi^{0,1}] = D_y[R_h\circ\chi_c]C_c^{-1}D_aT_\phi^{0,1}.$$

The operator norm is at most $\delta_{\text{chart}}\left\|C_c^{-1}D_aT_\phi^{0,1}\right\|_2$. Applying $\left\|A^\top W_R A\right\|_2 \le \|W_R\|_2\,\|A\|_2^2$ proves Eq. 37. □

The two errors are independent: a small residual value need not imply a small pulled-back residual differential. Both quantities are needed to characterize approximate decoders. Under differentiability, exact parameterization corresponds to $\varepsilon_{\text{chart}}=\delta_{\text{chart}}=0$.

H.5 Co-area law

For a density $\rho(x\mid c)$ relative to dV_{M_h}, hard conditioning by a full-row-rank residual is defined by a co-area target (Xu et al., 2026). On an injective chart, with $G_c(y)=J_\chi(y,c)^\top M_h J_\chi(y,c)$ and $J_R=D_xR_h|_{\chi_c(y)}$, the intrinsic density is

$$\pi_Y^{\mathrm{coa}}(y\mid c)=\frac{\rho(\chi_c(y)\mid c)}{Z_c}\,\frac{\sqrt{\det G_c(y)}}{\sqrt{\det\big(J_R(M_h)^{-1}J_R^\top\big)}}. \tag{38}$$

where Z_c normalizes the density; the numerator is the tangent volume element and the denominator is the residual-normal co-area Jacobian in the same metric.

Proposition 10 (Coordinate invariance of the co-area target)**.** *Let $\chi:\mathcal{Y}\to\mathcal{M}^\chi$ satisfy the hypotheses of Eq. 38, let $\varphi:\widetilde{\mathcal{Y}}\to\mathcal{Y}$ be a C^1 diffeomorphism, and set $\widetilde{\chi}=\chi\circ\varphi$. If π_Y^{coa} is the density in Eq. 38, then the density computed from $\widetilde{\chi}$ obeys*

$$\pi_{\widetilde{Y}}^{\mathrm{coa}}(\widetilde{y})=\pi_Y^{\mathrm{coa}}(\varphi(\widetilde{y}))\,|\det D\varphi(\widetilde{y})|\,, \tag{39}$$

The pushforward measures coincide: $\widetilde{\chi}_\#(\pi_{\widetilde{Y}}^{\mathrm{coa}}\,d\widetilde{y})=\chi_\#(\pi_Y^{\mathrm{coa}}\,dy)$. Reparameterization changes coordinate density but preserves the induced physical measure on the feasible component.

Proof. The chain rule gives $J_{\widetilde{\chi}}=J_\chi D\varphi$ and yields

$$\det(J_{\widetilde{\chi}}^\top M_x J_{\widetilde{\chi}})=\det(D\varphi)^2\det(J_\chi^\top M_x J_\chi).$$

The ambient density and residual-normal determinant are evaluated at the same physical point $\chi(\varphi(\widetilde{y}))$. Substitution into Eq. 38 yields Eq. 39; the ordinary change-of-variables theorem then proves equality of the pushforward measures. □

H.6 Encoding-induced representation error

For data outside the parameterization range, assume that the minimum below is attained and choose a measurable encoder selection

$$\eta_c(x) \in \arg\min_{y \in \mathcal{Y}_c} \|\chi_c(y) - x\|_{M_h}^2 . \tag{40}$$

For the encoder in Eq. 40, let $X \sim p_{\mathrm{data}}(\cdot \mid c)$ and define the feasible projection $\bar{X} = \chi_c(\eta_c(X))$. The joint law of $(X, \bar{X})$ is a valid coupling of p_{data} and $\bar{p}_{\mathrm{data}} = (\chi_c \circ \eta_c)_{\#} p_{\mathrm{data}}$. By the definition of the Wasserstein distance,

$$W_{2,M_x}^2(\bar{p}_{\mathrm{data}}, p_{\mathrm{data}}) \le \mathbb{E} \left\| \bar{X} - X \right\|_{M_x}^2 . \tag{41}$$

Eq. 41 bounds the representation error; we track it separately from generative error.

I Support, Measure, and Transport Errors

Residual factorization fixes feasible support, and the target-law construction specifies probability mass; matching determines how closely the learned pushforward follows that law. The following bound separates transport, measure, and representation error.

Proposition 11 (Three-way error bookkeeping)**.** *Let $\widehat{\pi}$, π_{enc}, π_{law}, and π_{data} be probability measures with finite second moments under the metric induced by M_x. Then*

$$W_{2,M_x}(\widehat{\pi}, \pi_{\mathrm{data}}) \le W_{2,M_x}(\widehat{\pi}, \pi_{\mathrm{enc}}) + W_{2,M_x}(\pi_{\mathrm{enc}}, \pi_{\mathrm{law}}) + W_{2,M_x}(\pi_{\mathrm{law}}, \pi_{\mathrm{data}}). \tag{42}$$

Proof. Apply the triangle inequality for W_{2,M_x} first through π_{enc} and then through π_{law}. □

Eq. 42 separates transport, target-law, and representation error. The latter two vanish for exactly feasible sample-defined targets; density-defined targets use the co-area law in Eq. 38.

J Physical and Statistical Preconditioning

After residual factorization, C_c rescales the decoder-induced physical metric in Eq. 6, while $P_{s,c}$ conditions the interpolation statistics seen by the network. We analyze their effects on local regression conditioning separately. Writing $J_{\Psi_c} = D_r \Psi_c$ and $y = \mu_c + C_c^{-1} r$, the chain rule gives $J_{\Psi_c}(r) = J_\chi(y, c) C_c^{-1}$, and the transformed metric is $\widetilde{G}_c(r) = J_{\Psi_c}(r)^\top M_h J_{\Psi_c}(r)$.

J.1 Regularized input covariance

Lemma 12 (Spectrum of regularized whitening)**.** *Let $C \succeq 0$ and $P = (C + \varepsilon I)^{-1/2}$ with $\varepsilon > 0$. If $C = U \operatorname{diag}(\lambda_i) U^\top$, then*

$$PCP^\top = U \operatorname{diag}\left(\frac{\lambda_i}{\lambda_i + \varepsilon} \right) U^\top . \tag{43}$$

On the active subspace of C,

$$\kappa_+(PCP^\top) = \frac{\lambda_{\max}(C)/(\lambda_{\max}(C) + \varepsilon)}{\lambda_{\min}^+(C)/(\lambda_{\min}^+(C) + \varepsilon)} . \tag{44}$$

Exact whitening is recovered when $\varepsilon = 0$ and C is nonsingular.

Proof. Applying $(\lambda + \varepsilon)^{-1/2}$ to the eigendecomposition of C gives Eq. 43 and the condition number in Eq. 44. □

For the interpolation states in Eq. 7, the transform in Eq. 8 gives

$$P_{s,c} \Sigma_{s,c} P_{s,c}^\top = \Sigma_{s,c} (\Sigma_{s,c} + \varepsilon_P I)^{-1} . \tag{45}$$

Eq. 45 shows how $P_{s,c}$ conditions network inputs while preserving the specified physical target law.

J.2 Local physical-space conditioning

Fix $s < t$, c, and a stopped-gradient target branch. Let $\zeta = P_{s,c}(r_s - m_{s,c})$ and consider a local linear velocity model $\bar{u}_A = A\zeta$. The endpoint depending on A is $a_A = r_s + (t-s)A\zeta$.

Proposition 13 (Local decoded-endpoint conditioning)**.** *The Gauss–Newton matrix of the decoded-endpoint loss with respect to A is*

$$H_A^{\mathrm{GN}} = \frac{t-s}{m\delta}\,\mathbb{E}\left[(\zeta\zeta^\top)\otimes \widetilde{G}_c(a_A)\right]. \tag{46}$$

Assume $\alpha I \preceq \widetilde{G}_c(a_A) \preceq \beta I$ in the stated region and let $S_{s,c} = \mathbb{E}[\zeta\zeta^\top]$ be positive definite. Then

$$\kappa\big(H_A^{\mathrm{GN}}\big) \le \frac{\beta}{\alpha}\kappa(S_{s,c}), \qquad S_{s,c} = P_{s,c}\Sigma_{s,c}P_{s,c}^\top \tag{47}$$

under exact centering.

Proof. Differentiating the squared decoded endpoint error gives the pullback and input factors in Eq. 46. The stated spectral bounds sandwich the expectation between positive multiples of $S_{s,c}\otimes I$. The extremal eigenvalues give Eq. 47. □

K One-Step Pullback Geometry

The pullback matrix $G_c(y) = J_\chi(y,c)^\top M_x J_\chi(y,c)$ is exact for infinitesimal perturbations. The action in preconditioned coordinates can be computed without forming a dense Jacobian:

$$\widetilde{G}_c(r)v = J_{\Psi_c}(r)^\top\big[M_h\big(J_{\Psi_c}(r)v\big)\big]. \tag{48}$$

Eq. 48 uses a decoder Jacobian-vector product followed by a vector-Jacobian product. The following result quantifies the approximation error for a finite coordinate displacement.

Proposition 14 (Finite-displacement pullback error)**.** *Fix y and a displacement d such that the segment $y + sd$, $s \in [0,1]$, lies in $\mathcal{Y}_c$. Freeze the positive-definite physical metric M_x. Suppose*

$$\big\|D_y^2\chi_c(y+sd)[v,v]\big\|_{M_x} \le L_\chi \|v\|_2^2 \tag{49}$$

for every $s \in [0,1]$ and v. Let $K_\chi = \|J_\chi(y,c)\|_{2\to M_x}$. Then

$$\big|\,\|\chi_c(y+d)-\chi_c(y)\|_{M_x}^2 - d^\top G_c(y)d\big| \le L_\chi K_\chi \|d\|_2^3 + \frac{L_\chi^2}{4}\|d\|_2^4. \tag{50}$$

Proof. Taylor's theorem under the bound in Eq. 49 gives

$$\chi_c(y+d)-\chi_c(y) = J_\chi(y,c)d + r_d, \qquad \|r_d\|_{M_x} \le \frac{L_\chi}{2}\|d\|_2^2.$$

Expanding the squared norm and applying Cauchy–Schwarz yields

$$2\,\|J_\chi d\|_{M_x}\|r_d\|_{M_x} + \|r_d\|_{M_x}^2,$$

which is bounded by the right-hand side of Eq. 50. □

The cubic remainder shows that the pullback metric is a local approximation, whereas decoded endpoint error measures finite physical discrepancy directly.

L Feasible Parameterizations and Preconditioning

L.1 Linear null-space parameterization

Proposition 15 (Exact encoding for affine constraints)**.** *Let $R_h(x) = Ax - b$, choose x_p with $Ax_p = b$, and let the columns of N form a basis of* $\ker A$*. Then*

$$\chi(y) = x_p + Ny \tag{51}$$

is an exact global constraint-satisfying parameterization of the affine feasible set. Evaluating the constraint yields

$$R_h(\chi(y)) = 0, \qquad J_R J_\chi = AN = 0. \tag{52}$$

Proof. Substituting Eq. 51 into $Ax - b$ gives Eq. 52. □

This construction covers discrete conservation, boundary, and observation constraints whenever they can be written as a compatible affine system. Numerical quality depends on the representation of the null-space basis; the exact algebraic identity and coordinate conditioning are separate properties.

L.2 Discrete streamfunction parameterization

Let D_x and D_y be discrete derivative matrices on the same grid and with the same boundary convention. Define

$$\chi(\psi) = \begin{bmatrix} D_y\psi \\ -D_x\psi \end{bmatrix}. \tag{53}$$

Proposition 16 (Compatible discrete incompressibility)**.** *If $D_xD_y = D_yD_x$, then every velocity decoded by Eq. 53 satisfies the matching discrete divergence exactly:*

$$D_xu + D_yv = (D_xD_y - D_yD_x)\psi = 0. \tag{54}$$

For the Euclidean state metric, the pullback matrix is

$$G_\psi = D_y^\top D_y + D_x^\top D_x. \tag{55}$$

Proof. Commutation gives the residual identity in Eq. 54. Eq. 55 follows by multiplying the Jacobian of Eq. 53 by the corresponding transpose factors. □

Eq. 55 separates exact incompressibility from intrinsic metric geometry: the parameterization enforces the former, while G_ψ determines the latter. On a periodic grid, the Fourier symbol of the pullback matrix scales like the squared discrete wave number. The pullback metric makes the physical importance of gradients and small scales explicit.

M Physical-Space Endpoint Matching

The two-time map in Eq. 9 is supervised at $s = t$ by the velocity loss in Eq. 10. At $s < t$, the loss in Eq. 12 compares two estimates of the same decoded physical endpoint. The following result links the learned finite-interval map to endpoint distribution error.

Let Y_s follow a target marginal path with $\dot{Y}_s = v_s(Y_s, c)$. Let

$$p_{\mathrm{enc},c} = \mathcal{L}(\Psi_c(Y_1))$$

denote the encoded target endpoint law. For the learned map, define the coordinate and physical defects

$$e_\theta(s,r) = \partial_s T_\theta^{s,1}(r,c) + D_r T_\theta^{s,1}(r,c)v_s(r,c), \qquad d_\theta(s,r) = J_{\Psi_c}\Big(T_\theta^{s,1}(r,c)\Big)\, e_\theta(s,r). \tag{56}$$

Proposition 17 (Physical flow-map defect bound)**.** *Assume a smooth target path, a valid coordinate domain, finite second moments, and*

$$\mathbb{E}\int_0^1 \|d_\theta(s,Y_s)\|_{M_x}^2\, ds < \infty.$$

For the defects in Eq. 56, the endpoint law $\widehat{p}_c$ of $\Psi_c(T_\theta^{0,1}(Y_0,c))$ satisfies

$$W_{2,M_x}^2(\widehat{p}_c, p_{\mathrm{enc},c}) \le \int_0^1 \mathbb{E}\left[e_\theta(s,Y_s)^\top \widetilde{G}_c\Big(T_\theta^{s,1}(Y_s,c)\Big)\, e_\theta(s,Y_s)\right] ds. \tag{57}$$

Proof. Along the target path,

$$\Psi_c(Y_1) - \Psi_c\Big(T_\theta^{0,1}(Y_0,c)\Big) = \int_0^1 d_\theta(s,Y_s)\, ds.$$

The common source Y_0 couples the endpoint laws; Cauchy–Schwarz and $\widetilde{G}_c$ give Eq. 57. □

The consistency loss is a finite-difference surrogate for the flow-map defect controlled by Eq. 57. For the interpolation path, let

$$v_s(r, c) = \mathbb{E}[w \mid r_s = r, c].$$

For fixed s, t, c, δ and a stopped-gradient target, conditioning on r_s and averaging over the endpoint-pair velocity w gives the corresponding conditional semi-gradient with respect to the differentiable velocity output:

$$-\frac{1}{m}\widetilde{G}_c(a_\theta)\left(\partial_s T_\theta^{s,t} + D_r T_\theta^{s,t} v_s\right) + O(\delta). \tag{58}$$

Eq. 58 fixes the target branch and uses δ to normalize the finite difference.

N Regular Residual Geometries

For regular residual manifolds, PMosFM constructs local feasible coordinates from discrete residual-zero sets rather than assuming a prescribed Riemannian manifold (Chen & Lipman, 2024).

N.1 Local residual-manifold parameterizations

Fix the mesh h and condition c, and let $M_x = M_h$ be positive definite. Assume that $R_h(\cdot, c)$ is C^2 on an open neighborhood and that $D_x R_h$ has constant rank q there. On the regular component

$$\mathcal{M}_c = \{x : R_h(x, c) = 0\}, \qquad m = d - q,$$

the tangent space is $T_x\mathcal{M}_c = \ker D_x R_h(x, c)$, with physical metric $g_x(\xi, \zeta) = \xi^\top M_x \zeta$.

Proposition 18 (Local feasible parameterizations)**.** *Every point of a regular component has a local injective parameterization $\chi_c : \mathcal{Y}_c \to \mathcal{M}_c$ with $R_h \circ \chi_c = 0$. For a compact regular target region, finitely many such parameterizations cover the region. After an invertible coordinate transform $C_{j,c}$, each decoder*

$$\Psi_{j,c}(r) = \chi_{j,c}(\mu_{j,c} + C_{j,c}^{-1} r)$$

remains exactly feasible and has continuous pullback metric $\widetilde{G}_{j,c}(r) = D\Psi_{j,c}(r)^\top M_x D\Psi_{j,c}(r)$.

A near-identity metric bound for $C_{j,c}$ requires separate verification; the construction is local.

The constructions above apply patchwise. Let $\Sigma_{s,j,c}$ denote the covariance of r_s within patch j, and define

$$P_{s,j,c} = (\Sigma_{s,j,c} + \varepsilon_P I)^{-1/2}, \qquad \widetilde{G}_{j,c}(r) = D\Psi_{j,c}(r)^\top M_x D\Psi_{j,c}(r).$$

Under the assumptions of Proposition 13, if $\alpha_{j,c} I \preceq \widetilde{G}_{j,c}(r) \preceq \beta_{j,c} I$ on the evaluated region, then

$$\kappa(H_A^{\mathrm{GN}}) \le \frac{\beta_{j,c}}{\alpha_{j,c}} \kappa\left(P_{s,j,c}\Sigma_{s,j,c}P_{s,j,c}^\top\right). \tag{59}$$

For overlapping patches, let a_j form a partition of unity on the target support and set $\omega_j = \int a_j \, d\nu_c$, so $\sum_j \omega_j = 1$. For $\omega_j > 0$, define the normalized patch law

$$\nu_{j,c} = \frac{a_j \nu_c}{\omega_j},$$

and let $\widehat{\nu}_{j,c}$ denote the corresponding generated endpoint law. Set

$$\widehat{\nu}_c = \sum_j \omega_j \widehat{\nu}_{j,c}.$$

Applying Proposition 17 within each patch and mixing the resulting couplings gives

$$W_{2,M_x}^2(\widehat{\nu}_c, \nu_c) \le \sum_j \omega_j \mathcal{E}_{j,c}, \tag{60}$$

where $\mathcal{E}_{j,c}$ is the integrated physical flow-map defect from Eq. 57, evaluated on patch j. Eqs. 59 and 60 extend the conditioning and endpoint-error bounds locally without a global flattening of the residual manifold.

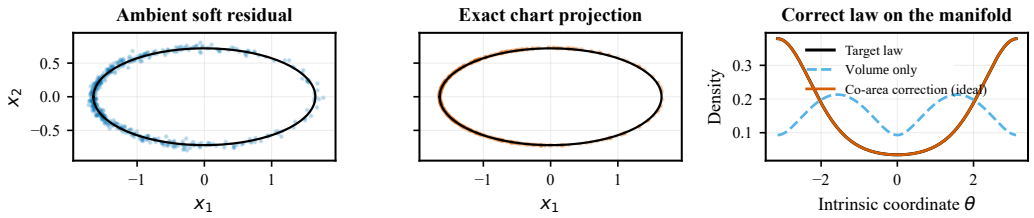


Figure 16: Support and measure on a residual manifold. Exact parameterization enforces feasibility, and the co-area correction recovers the target law.

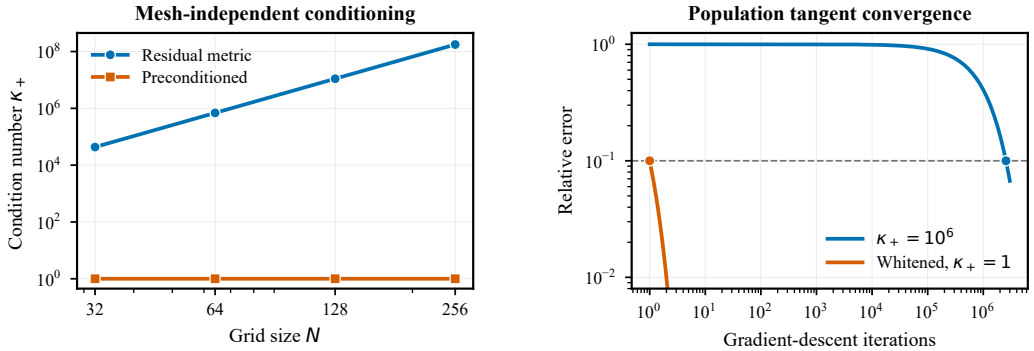


Figure 17: Normal and tangent conditioning. Left: preconditioning controls residual-normal conditioning under grid refinement. Right: whitening accelerates tangent-regression convergence.

O Additional Analyses

Supplementary evaluations follow the dataset partitions, units, and evaluators in Appendix C.

O.1 Feasibility, Distribution, and Transport

Appendices D.4, D.5, and D.6 describe support, decoder, and conditioning diagnostics. Fig. 16 shows that exact parameterization places samples on the feasible set, while the volume-only law still differs from the target density; the co-area correction reduces this mismatch.

Table 18: Normal conditioning.

N	$\kappa(A_h^\top A_h)$	$\kappa(A_h^\top W_h A_h)$
32	4.33×10^4	1.0001
64	6.90×10^5	1.0001
128	1.10×10^7	1.0001
256	1.76×10^8	1.0001

O.2 Decoder tolerance and cost

Decoder formulation. For a source coordinate r_0, PMosFM first evaluates the learned finite-interval map and then applies the geometric inverse transform:

$$\widehat{r}_1 = T_\theta^{0,1}(r_0, c), \qquad \widehat{y}_1 = \mu_c + C_c^{-1}\widehat{r}_1, \qquad \widehat{x}_1^{(K)} = \chi_c^{(K)}(\widehat{y}_1). \tag{61}$$

In Eq. 61, K indexes the physical decoder, not the learned transport map; for an explicit constraint-satisfying parameterization, decoding applies the algebraic or compatible discrete differential operations defining χ_c, covering the affine, algebraic, and compatible potential-based cases considered in the experiments:

$$\widehat{x}_1 = \chi_c(\widehat{y}_1), \qquad R_h(\chi_c(y), c) \equiv 0,$$

For an implicit parameterization, write $x = (y, z)$. Here $z^\star(y, c)$ denotes the selected solution branch, whereas $\mathcal{S}_h$ denotes one decoder iteration:

$$F_h(z^\star, y, c) = 0, \qquad z^{(k+1)} = \mathcal{S}_h(z^{(k)}; y, c), \qquad \chi_c^{(K)}(y) = \begin{bmatrix} y \\ z^{(K)}(y, c) \end{bmatrix}. \tag{62}$$

The Darcy implementation uses the decoder in Eq. 62, with $K = 256$ in this configuration. For an exact implicit case $F_h(z, y, c) = 0$ with invertible $D_z F_h$, the directional derivative satisfies

$$D_y z[v] = -(D_z F_h)^{-1} D_y F_h[v].$$

Finite-iteration decoders require separate derivative-accuracy checks.

Decoder accuracy and residual bounds. For a K-iteration decoder, define the value and differential diagnostics

$$\begin{aligned} \varepsilon_{\mathrm{dec}}^{(K)} &:= \sup_{y \in \mathcal{Y}_c} \left\| R_h(\chi_c^{(K)}(y), c) \right\|_2, \\ \delta_{\mathrm{dec}}^{(K)} &:= \sup_{y \in \mathcal{Y}_c} \left\| D_y \left[R_h \circ \chi_c^{(K)} \right] (y, c) \right\|_2. \end{aligned} \tag{63}$$

In Eq. 63, the first diagnostic measures physical residual, and the second measures its differential in the represented coordinates. With

$$G_{\theta,c}^{(K)} = \chi_c^{(K)} \circ \left(\mu_c + C_c^{-1} \cdot \right) \circ T_\theta^{0,1},$$

the residual bound gives

$$\left\| R_h \left(G_{\theta,c}^{(K)}(r_0, c), c \right) \right\|_2 \le \varepsilon_{\mathrm{dec}}^{(K)}.$$

For any differentiable source or model variable a,

$$\left\| D_a \left[R_h \circ G_{\theta,c}^{(K)} \right] \right\|_2 \le \delta_{\mathrm{dec}}^{(K)} \left\| C_c^{-1} D_a T_\theta^{0,1} \right\|_2.$$

The two diagnostics address different properties: residual magnitude and the residual-normal differential after approximate decoding. The corresponding residual Gauss–Newton block satisfies

$$\begin{aligned} H_{\mathrm{GN},a}^{R,(K)} &:= \left(D_a [R_h \circ G_{\theta,c}^{(K)}] \right)^\top W_R \, D_a [R_h \circ G_{\theta,c}^{(K)}], \\ \left\| H_{\mathrm{GN},a}^{R,(K)} \right\|_2 &\le \|W_R\|_2 \left(\delta_{\mathrm{dec}}^{(K)} \right)^2 \left\| C_c^{-1} D_a T_\theta^{0,1} \right\|_2^2. \end{aligned} \tag{64}$$

For an exact decoder, $\varepsilon_{\mathrm{dec}} = \delta_{\mathrm{dec}} = 0$ and the encoded residual-normal block vanishes; Eq. 64 bounds the block retained by an iterative decoder.

End-to-end cost. We distinguish learned transport from physical decoding:

$$T_{\mathrm{sample}}^{(K)} = T_{\mathrm{map}} + T_{\mathrm{pre}} + T_{\mathrm{dec}}^{(K)}, \qquad \mathrm{NFE}_{\mathrm{learned}} = 1, \tag{65}$$

In Eq. 65, T_{map} is the cost of $T_\theta^{0,1}$, T_{pre} contains the fixed coordinate and input transforms, and $T_{\mathrm{dec}}^{(K)}$ is the physical decoding cost. For a direct setting $T_{\mathrm{dec}}^{\mathrm{dir}} = T_\chi$; for an iterative decoder,

$$T_{\mathrm{dec}}^{(K)} = T_{\mathrm{init}} + \sum_{k=0}^{K-1} T_{\mathcal{S}}^{(k)} + T_{\mathrm{post}} \approx T_{\mathrm{init}} + K \, \overline{T}_{\mathcal{S}} + T_{\mathrm{post}}.$$

The timing studies report learned transport and full physical decoding separately, while the residual diagnostics quantify the accuracy of the realized PMosFM endpoint.