STRUCTURE-PRESERVING FROZEN NEURAL OPERATORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Coupled PDE solvers must enforce physical constraints while controlling optimization error and saddle-point conditioning. We introduce Structure-Preserving Frozen Neural Operators (SP-FNO), a frozen Fourier representation with coefficient-space constraints and gradient-free solves. A rank-revealing SVD yields an orthogonal projector for linear constraints. For mixed systems, we solve the velocity problem on the constraint kernel and recover pressure through a Schur complement governed by the retained inf-sup spectrum, independent of the constraint-block coupling scale. On boxes, analytic Gram assembly yields quadrature-free functional error majorants, valid with proven constants in exact arithmetic; our float64 evaluations do not enclose roundoff. The gradient-free core attains relative L^2 errors of 2.14×10^{-9}, 1.50×10^{-9}, and 1.90×10^{-4} on manufactured Stokes, Maxwell, and Navier–Stokes problems without frequency anchoring. For advection whose solution lies in the frozen span, it reaches 2.83×10^{-14} error in 1.19×10^{-2}s. A separate learned field-space hybrid achieves best mean VRMSE on all five The Well datasets tested under the $4 \rightarrow 1$ protocol, reaching 0.01668 one-step VRMSE on active matter, using a 17.34M-parameter backbone and an eight-pass symmetry ensemble. Code: `https://anonymous.4open.science/r/SP-FNO-C44A`.

1 INTRODUCTION

Neural operators learn PDE solution maps from data (Li et al., 2021), while physics-informed neural networks fit solutions through differential residuals (Raissi et al., 2019). Both require optimization; multi-term physics losses can be difficult to train (Krishnapriyan et al., 2021; Wang et al., 2024). Frozen-feature methods replace network training with coefficient solves (Datar et al., 2026; Sulc, 2026). For coupled PDEs, this leaves the choice of constraint representation, mixed-system stability, and error estimation.

In incompressible flow and electromagnetism, low residual error alone does not ensure divergence constraints, boundary traces, or stable pressure recovery. Classical null-space and mixed methods address these issues through compatible spaces (Benzi et al., 2005; Boffi et al., 2013). We adapt this principle to frozen Fourier features: represent a linear constraint by $D\mathbf{c} = 0$, construct its kernel with a rank-revealing SVD, and solve within that kernel. For sampled constraints, this enforces the discrete equations; continuum preservation requires a conforming basis or an exact representation of the constraint operator.

For saddle-point problems, we separate a positive-definite velocity solve on $\ker(B_r)$ from pressure recovery. The Schur condition number depends on the retained inf-sup spectrum and, for a fixed energy matrix, is independent of the constraint-block scale. On boxes, analytic assembly also yields quadrature-free functional majorants. Their validity requires proven constants and exact arithmetic; the float64 evaluations do not enclose roundoff.

The frozen core uses linear least squares, analytic Newton updates, or constrained Gram dynamics for time-dependent problems. For richer fields, we use a learned hybrid that projects field-space corrections onto the physical constraints (Figure 1). Both routes impose the constraint through the representation, without penalty terms.

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.

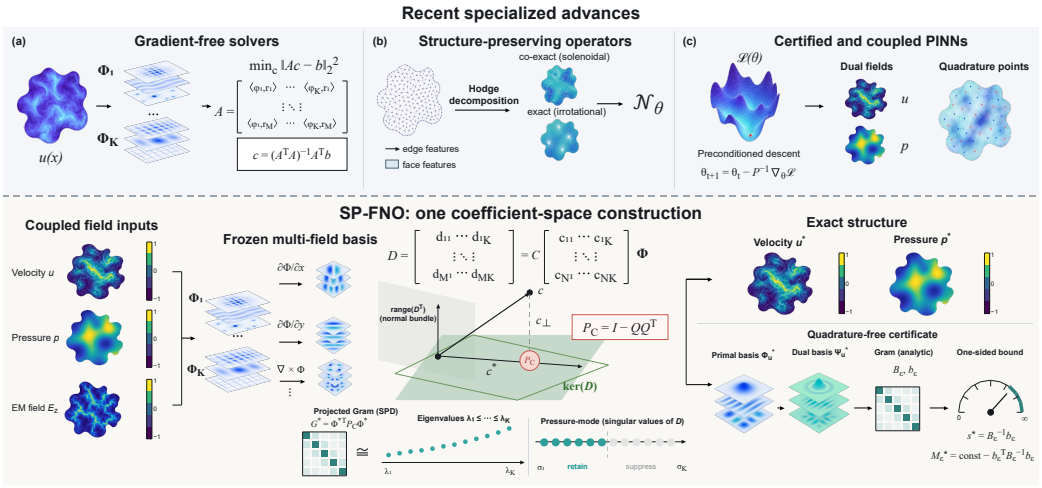

Figure 1: SP-FNO at a glance. Top: related approaches. Bottom: the SP-FNO components.

We evaluate representability, constraint preservation, mixed-system stability, and majorant effectivity on advection and manufactured PDEs. Active-matter experiments compare one-step and rollout errors while accounting for model size and inference cost.

2 Related Work

Frozen features and least-squares PDE solvers. Rahimi & Recht (2007) introduced random Fourier features to approximate shift-invariant kernels with finite feature maps and linear solvers. For PDEs, Dong & Li (2021) combined fixed hidden layers with domain decomposition and linear or nonlinear least squares; Chen et al. (2022) developed collocation with multiscale random features and loss rescaling. Frozen-PINN (Datar et al., 2026) and FastLSQ (Sulc, 2026) continue this coefficient-fitting approach, while frequency-adaptive variants improve high-frequency resolution (Ren et al., 2025; Xiong et al., 2026). Preconditioned physics losses address sensitivity to coupling strength (Park et al., 2026). We build on fixed-feature approximation by eliminating linear constraints before solving and deriving the resulting mixed-system conditioning.

Constraints and mixed formulations. Benzi et al. (2005) review null-space and Schur-complement methods for saddle-point systems, while mixed finite-element theory relates stability to compatibility between primal and multiplier spaces (Boffi et al., 2013). Null-space treatment of Neumann problems (Bochev & Lehoucq, 2005) and divergence-free discretizations (Cockburn et al., 2007) provide further precedents for imposing constraints through the representation. Neural approaches include divergence-free linearized networks (He et al., 2026) and equality-constrained incompressible training (Hu & Senocak, 2025). We apply these principles to frozen Fourier coefficients with a rank-revealing SVD. The resulting pressure solve retains dependence on the selected inf-sup spectrum; removing weak modes can improve conditioning at the cost of pressure resolution.

Functional error estimates. Repin (2008) develops functional a posteriori estimates that bound the error of an admissible approximation through residuals, dual fields, and domain constants, without requiring a particular discretization. Astral (Fanaskov et al., 2024) uses functional majorants to train PINNs, and neural verification methods study their reliable evaluation (Ernst et al., 2025). We use this framework to derive closed-form Gram and load matrices for supported Fourier spaces on boxes. Analytic assembly removes quadrature error and reduces dual-field optimization to a coefficient solve; validated numerical bounds also require control of floating-point error.

Learned operators and refinement. DeepONet encodes input functions and output coordinates through branch and trunk networks (Lu et al., 2021); FNO learns spectral transformations (Li et al., 2021). Kovachki et al. (2023) formulate neural operators between function spaces and establish approximation results. Flowers models transport through learned warps (Muser et al., 2026), IRNO

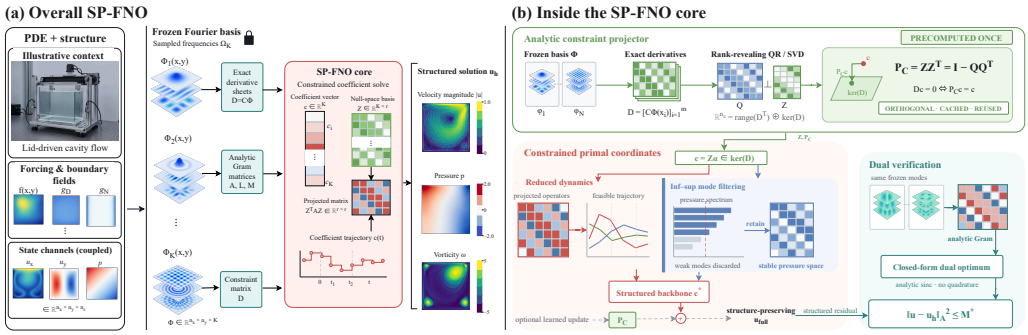

Figure 2: SP-FNO architecture. Left: the pipeline from PDE plus structure to the structured solution. Right: inside the core, deriving $D = C\Phi$, obtaining the null-space basis Z, assembling the analytic matrices on $\ker(D)$, and adding an optional learned residual projected by $P_{\mathcal{C}}$, preserving the constraint represented by D or the field-space projector.

refines pretrained operator predictions (Liu et al., 2026), and WaveLiT uses architectural priors for parameter efficiency (Sankaran et al., 2026). Our frozen core solves new right-hand sides in a shared constrained span. The learned hybrid instead shares trained weights across inputs and projects each field-space correction onto the physical constraints.

3 Method

SP-FNO combines a frozen feature representation, linear constraint elimination, and analytic residual assembly (Figure 2). The coefficient solver requires no gradient descent; Section 3.6 describes the separate learned hybrid used for active matter.

3.1 Frozen representation and problem setup

We consider a vector PDE with a linear structure constraint,

$$\mathcal{L}\mathbf{u} = \mathbf{f} \text{ in } \Omega, \qquad \mathcal{B}\mathbf{u} = \mathbf{g} \text{ on } \partial\Omega, \qquad \mathcal{C}\mathbf{u} = 0, \tag{1}$$

and write $\mathbf{u}_h(x) = \Phi(x)\mathbf{c}$, where Φ collects component-wise features

$$\phi_j(x) = \sin(\omega_j^\top x + b_j), \qquad \omega_j \sim p(\omega), \quad b_j \sim \mathrm{Unif}[0, 2\pi]. \tag{2}$$

We sample the frequencies once and solve for $\mathbf{c}$ by constrained least squares, a reduced linear system, or Newton iteration. Derivatives satisfy $D^\alpha \phi_j = \omega_j^\alpha \sin(\omega_j \cdot x + b_j + |\alpha|\pi/2)$. For K field components and N features per component, $\mathbf{c} \in \mathbb{R}^{KN}$. For fixed PDE coefficients and geometry, the core defines the input-to-solution map $(\mathbf{f}, \mathbf{g}) \mapsto \Phi\mathbf{c}^\star(\mathbf{f}, \mathbf{g})$. For fixed geometry, operator, and basis, this map reuses the constraint factorization and applicable Gram matrices; each input requires a new coefficient solve.

On boxes, complex-exponential identities give closed-form feature and derivative Gram matrices (Appendix F). Analytic load assembly also requires compatible forcing. Our manufactured problems use continuous frequency distributions; periodic advection uses integer sine/cosine pairs.

3.2 Constraint projection and admissible fields

Substitution into the linear constraint gives $D\mathbf{c} = 0$. The rows of D encode the chosen constraints, such as sampled divergence or boundary traces. Figure 3 illustrates the coefficient geometry.

Proposition 1 (Constraint projector)**.** *For a matrix D with rank-r SVD $D = U_r\Sigma_r V_r^\top$, the Euclidean orthogonal projector onto* $\ker(D)$ *is*

$$P_{\mathcal{C}} = I - D^\dagger D = I - V_r V_r^\top = ZZ^\top, \tag{3}$$

where Z is an orthonormal null-space basis. This holds for rectangular and rank-deficient matrices. With a truncated SVD, it projects onto the numerical null space at the chosen tolerance.

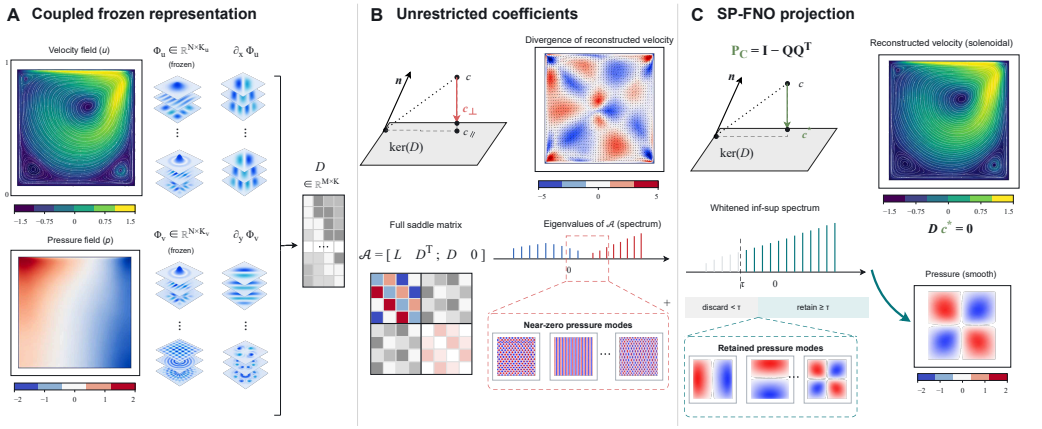

Figure 3: From unconstrained coefficients to structure. The coupled frozen representation and constraint matrix; the unrestricted coefficients off the constraint manifold; and the SP-FNO projection for constrained velocity solution and filtered pressure recovery.

We retain singular values above $10^{-8}\sigma_{\max}(D)$ and solve in coordinates $\mathbf{c} = Z\alpha$. The SVD costs $O(mn\min(m,n))$ for $D \in \mathbb{R}^{m\times n}$ and is reused across solves. The proof and treatment of inhomogeneous constraints appear in Appendix A.

Let D_τ retain the singular values above $\tau_{\text{svd}}\sigma_1(D)$ and let σ_{disc} be the largest discarded singular value, or zero if none is discarded. Then the implemented projector satisfies

$$D_\tau P_{\mathcal{C}} = 0, \qquad \|DP_{\mathcal{C}}\|_2 = \sigma_{\text{disc}} \le \tau_{\text{svd}}\sigma_1(D). \tag{4}$$

Thus the residual for projected coefficients is bounded by $\sigma_{\text{disc}}\|\mathbf{c}\|_2$. Exact feasibility for the original discrete operator requires retaining every nonzero singular direction; otherwise feasibility depends on the truncation tolerance and coefficient scale.

Collocation projection enforces divergence only at the sampled points. For continuum incompressibility in two dimensions, we use $\mathbf{v}_h = (-\partial_y\psi_h, \partial_x\psi_h)$. Choosing $\psi_h = x^2(1-x)^2y^2(1-y)^2\sum_j a_j\phi_j$ also enforces zero velocity trace on $[0,1]^2$. The polynomial envelope preserves analytic integration. These continuum conditions are needed for the Stokes bounds below.

3.3 Mixed systems and inf-sup filtering

For the mixed component-basis formulation of Stokes, let $A \succ 0$ be the velocity energy matrix, M_p the pressure mass matrix after a zero-mean gauge, and $B_{aj} = -\int_\Omega \psi_a \nabla\cdot\phi_j$. With $A = L_v L_v^\top$ and $M_p = L_p L_p^\top$, define

$$\widehat{B} = L_p^{-1} B L_v^{-\top}, \qquad T_p = L_p^{-\top} U_r, \qquad B_r = T_p^\top B, \tag{5}$$

where U_r contains the left singular vectors of $\widehat{B}$ with $\sigma_i > \tau\sigma_1$. Thus $T_p^\top M_p T_p = I$; filtering selects pressure directions in the mass/energy-normalized spectrum.

For the saddle system

$$\begin{pmatrix} A & \gamma B_r^\top \\ \gamma B_r & 0 \end{pmatrix}\begin{pmatrix} \mathbf{c_v} \\ \mathbf{p}_r \end{pmatrix} = \begin{pmatrix} \mathbf{f}_v \\ \mathbf{g}_r \end{pmatrix}, \tag{6}$$

write $\mathbf{c_v} = \mathbf{c}_{\text{part}} + Z\mathbf{c}_z$, with $B_r\mathbf{c}_{\text{part}} = \gamma^{-1}\mathbf{g}_r$ and Z an orthonormal basis of $\ker(B_r)$. The velocity and pressure solves are

$$(Z^\top AZ)\mathbf{c}_z = Z^\top(\mathbf{f}_v - A\mathbf{c}_{\text{part}}), \tag{7}$$

$$S\mathbf{p}_r = \gamma^{-1}B_rA^{-1}\mathbf{f}_v - \gamma^{-2}\mathbf{g}_r, \quad S = B_rA^{-1}B_r^\top. \tag{8}$$

Proposition 2 (Decoupled conditioning). *Assume $A \succ 0$, B_r has full row rank, and A, B_r do not depend on $\gamma > 0$. Then both solve matrices are SPD and*

$$\kappa_2(Z^\top AZ) \le \kappa_2(A), \qquad \kappa_2(S) = \left(\frac{\sigma_1(\widehat{B})}{\sigma_r(\widehat{B})}\right)^2 \le \tau^{-2}. \tag{9}$$

Their conditioning is independent of the constraint-block scale γ.

This result concerns coupling confined to the off-diagonal blocks of equation 6; it does not cover coupling-dependent constitutive matrices. Filtering controls stability within the retained pressure space and may discard approximation directions. Appendix E gives the threshold study; Algorithm 2 gives the solver, with the proof in Appendix A. The collocation variant uses a separate pressure-gradient filter without an LBB guarantee.

3.4 TIME-DEPENDENT PROBLEMS

For linear $\mathcal{L}$ in $\partial_t \mathbf{u} + \mathcal{L}\mathbf{u} = \mathbf{f}$, let $M = \int_\Omega \Phi^\top \Phi$, $L = \int_\Omega \Phi^\top \mathcal{L}\Phi$, and $\mathbf{f}_\Phi = \int_\Omega \Phi^\top \mathbf{f}$. The constrained Gram system is

$$(Z^\top M Z)\dot{\alpha} + Z^\top L Z \alpha = Z^\top \mathbf{f}_\Phi, \qquad \mathbf{c}(t) = Z\alpha(t). \tag{10}$$

A matrix exponential integrates the linear reduced dynamics, with an augmented system for piecewise-constant forcing. Nonlinear time evolution uses a semi-implicit step. The trajectory remains in the represented constraint space; time integration does not remove error from unresolved frequencies.

3.5 ANALYTICALLY ASSEMBLED ERROR MAJORANTS

For Stokes, define $\|e\|_{\mathcal{A}}^2 = \nu \|\nabla e\|_{L^2}^2$. The following functional estimate uses an admissible total-stress reconstruction.

Proposition 3 (Velocity energy majorant)**.** *Assume (D0) $\mathbf{u}_h \in H_0^1(\Omega)^d$ is divergence-free in the continuum; (D1) $\Sigma_h \in H(\mathrm{Div};\Omega)$, with row-wise divergence; and (D2) C_Ω is a proven Poincaré constant. For $\varepsilon > 0$,*

$$\|\mathbf{u} - \mathbf{u}_h\|_{\mathcal{A}}^2 \le \mathcal{M}_\varepsilon = \frac{1}{\nu}\left[(1+\varepsilon)\,\|\Sigma_h - \nu\nabla\mathbf{u}_h\|_{L^2}^2 + (1+\varepsilon^{-1})C_\Omega^2\,\|\nabla\cdot\Sigma_h + \mathbf{f}\|_{L^2}^2\right]. \tag{11}$$

Representing $\Sigma_h = \Sigma\mathbf{s}$ in the frozen feature family gives $\mathcal{M}_\varepsilon = \mathbf{s}^\top B_\varepsilon \mathbf{s} - 2\mathbf{s}^\top \mathbf{b}_\varepsilon + c_\varepsilon$ with analytic Gram blocks. For $B_\varepsilon \succ 0$, the minimizer is $\mathbf{s}^\star = B_\varepsilon^{-1}\mathbf{b}_\varepsilon$ and the bound is $c_\varepsilon - \mathbf{b}_\varepsilon^\top B_\varepsilon^{-1}\mathbf{b}_\varepsilon$. Pressure estimation additionally requires a proven continuous inf-sup constant β for the velocity test norm $\|\nabla\mathbf{v}\|_{L^2}$. For pressures in the same zero-mean gauge and momentum residual $r_h = \mathbf{f} + \nu\Delta\mathbf{u}_h - \nabla p_h$, we have

$$\|p - p_h\|_{L^2} \le \beta^{-1}\left[C_\Omega\,\|r_h\|_{L^2} + \nu\,\|\nabla(\mathbf{u} - \mathbf{u}_h)\|_{L^2}\right]. \tag{12}$$

Equation equation 11 bounds the second term by $\sqrt{\nu\mathcal{M}_\varepsilon}$. The discrete singular values used for pressure filtering do not establish β, and an independently chosen dual stress's equilibrium residual cannot replace r_h. Thus retained-space stability and a continuous pressure error bound require different evidence (Appendix B).

The Poisson and Maxwell experiments evaluate functional-dual majorants. The Stokes experiments use the separate strong-residual bound $\|e\|_{\mathcal{A}} \le C_\Omega \nu^{-1/2}\,\|\mathbf{f} - \mathcal{L}\mathbf{u}_h\|_{L^2}$ under the same primal admissibility conditions (Appendix F). These inequalities hold in exact arithmetic; float64 evaluation without roundoff enclosure yields *certificate-style majorants*, rather than validated numerical certificates.

3.6 PROJECTED LEARNED CORRECTIONS

A coefficient-space correction preserves the represented constraint through

$$\mathbf{u}^{\mathrm{full}}(x) = \Phi(x)\big[\mathbf{c}^\star + P_{\mathcal{C}} r_\theta(\mathbf{c}^\star)\big]. \tag{13}$$

This correction remains in the same feature span. The active-matter experiments instead use a learned grid operator with a field-space projection Π_{phys}: spectral Hodge projection for velocity, symmetric D with $\mathrm{tr}(D) = c$, and symmetric trace-free E. We denormalize channels before projection and renormalize afterward. Each learned refinement step ends with the same projection.

For symmetry ensembling we average inverse-transformed predictions over D_4. On normalized data, the group action is $\rho(g) = N \circ g_{\mathrm{phys}} \circ N^{-1}$, where N is channel normalization. Physical

Table 1: Periodic linear advection, $u_t + 40\,u_x = 0$ on $[0, 2\pi] \times [0, 1]$ with $u(x, 0) = \sin x$. Published rows follow the protocol of Datar et al. (2026); the Fourier and Galerkin rows are reproduced by us on the same test.

Method	Relative L^2 error ↓	Time (s) ↓
PINN (L-BFGS, 4 layers, 5,000 epochs)	$6.92 \times 10^{-1} \pm 2.96 \times 10^{-2}$	30.5
Causal PINN (2024)	2.90 ± 1.2	357.63
Frozen-PINN-elm (2026)	$4.19 \times 10^{-3} \pm 2.97 \times 10^{-3}$	0.67
Frozen-PINN-swim (2026)	$8.42 \times 10^{-9} \pm 1.12 \times 10^{-8}$	0.7
Mesh-based IGA (reported)	1.17×10^{-10}	0.07
Fourier spectral, 7 modes (ours)	2.09×10^{-15}	7×10^{-5}
Galerkin P_1, 256 cells (ours)	2.18×10^{-1}	1.8×10^{-2}
Galerkin P_1, 4096 cells (ours)	1.52×10^{-2}	4.1×10^{-1}
SP-FNO (ours)	$\mathbf{2.832 \times 10^{-14} \pm 4.76 \times 10^{-17}}$	$\mathbf{1.19 \times 10^{-2} \pm 3.79 \times 10^{-4}}$

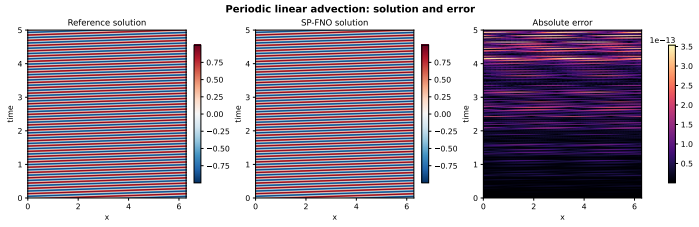

Figure 4: Periodic advection over the displayed time window: reference solution, SP-FNO solution, and absolute error.

rotations act on vector and tensor components as well as spatial coordinates. Algorithm 1 gives the prediction procedure; Appendix C specifies the channels and inference budgets. This learned hybrid shares the constraint-preservation principle of the frozen core; it does not inherit its gradient-free fitting or error-majorant guarantees.

Algorithm 1 Field-space SP-FNO hybrid for active matter

Require: Four normalized input states $\mathbf{x}$, backbone F_θ, normalization N, group $\mathcal{G} \subseteq D_4$, optional refiner r_ϕ

$\hat{\mathbf{y}}_0 \leftarrow |\mathcal{G}|^{-1} \sum_{g \in \mathcal{G}} \rho(g^{-1}) F_\theta(\rho(g)\mathbf{x})$
$\hat{\mathbf{y}}_{\text{phys}} \leftarrow \Pi_{\text{phys}}(N^{-1}(\hat{\mathbf{y}}_0))$
for $i = 1, \ldots, K$ **do**
    $\hat{\mathbf{y}}_{\text{phys}} \leftarrow \Pi_{\text{phys}}(\hat{\mathbf{y}}_{\text{phys}} + r_\phi(\hat{\mathbf{y}}_{\text{phys}}))$
end for
return $N(\hat{\mathbf{y}}_{\text{phys}})$

4 Experiments

We evaluate periodic advection, manufactured Stokes, Maxwell/TE and steady Navier–Stokes, and learned prediction on five The Well datasets (Ohana et al., 2024). We report relative L^2 error and residuals for PDE solves, and VRMSE for the learned field-space hybrid of Section 3.6. The latter uses a trained backbone; the manufactured PDE solver uses frozen features and linear algebra. Appendix D gives seeds, hardware, aggregation rules, and timing details.

Advection and representability. On the 256×100 float64 benchmark, SP-FNO reaches 2.832×10^{-14} error in 1.19×10^{-2}s (Table 1; Figure 4). The seven integer Fourier pairs contain the exact travelling wave, so this tests linear-algebra accuracy. The classical Fourier solve is more accurate and faster on this matched basis. With 16 pairs, however, $\sin(9x) + 0.5\sin(17x)$ incurs $\sqrt{0.2} = 4.47 \times 10^{-1}$ error despite an 8×10^{-13} strong residual: the basis omits the 17th harmonic (Table 23). The broad-spectrum wave packet has 3.957×10^{-3} error with the same basis. For representable initial

Table 2: SP-FNO on five The Well datasets under the $4 \to 1$ protocol (mean VRMSE). Baselines follow Muser et al. (2026); – denotes no published result. WaveLiT comparisons appear in Table 7.

	One-step ↓					T[1:20] Rollout ↓					T[21:60] Rollout ↓				
Dataset	FNO	CNextU-Net	scOT	Flower	**SP-FNO**	FNO	CNextU-Net	scOT	Flower	**SP-FNO**	FNO	CNextU-Net	scOT	Flower	**SP-FNO**
Active Matter	0.1749	0.0650	0.1050	0.0249	**0.01668**	3.2862	1.7781	4.1055	1.3905	**0.75143**	1.8214	1.6465	2.0372	1.5251	**1.39739**
MHD 64	0.3403	0.2062	–	0.1165	**0.0951**	1.3007	0.9534	–	0.7580	**0.6841**	1.6366	1.4299	–	1.2827	**1.0558**
Neutron Star	0.4452	0.3391	–	0.3269	**0.2841**	0.5980	0.6529	–	0.6223	**0.5462**	0.6209	0.7938	–	0.7202	**0.5796**
TRL 2D	0.3250	0.2559	0.3555	0.1930	**0.1867**	1.2328	0.7051	0.8299	0.5491	**0.4507**	6.0891	1.3390	1.7280	1.0516	**0.9529**
TRL 3D	0.3261	0.3322	–	0.2073	**0.1452**	0.9203	0.7139	–	0.6840	**0.6018**	1.9969	0.7885	–	0.9881	**0.6879**

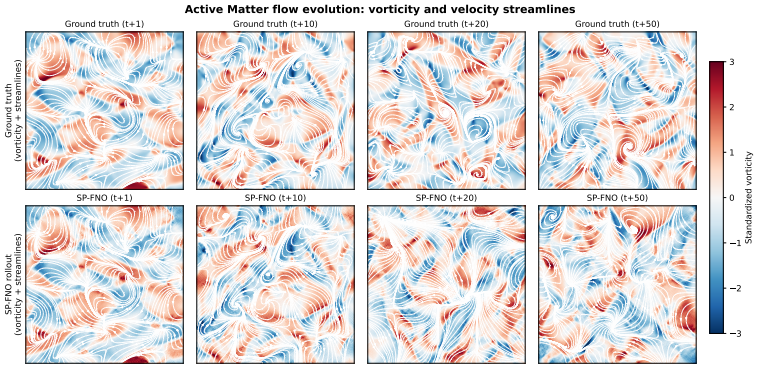

Figure 5: Active-matter flow evolution on the The Well test set, ground truth above and SP-FNO rollout below.

Table 3: SP-FNO on manufactured coupled PDEs without a frequency prior: Stokes with $\nu = 0.01$, Maxwell/TE with $\kappa = 3$, and steady Navier–Stokes (NS) with analytic Newton iteration.

Task	Rel. L^2 (u) ↓	Rel. L^2 (v) ↓	Rel. L^2 (p) ↓
Stokes	2.139×10^{-9}	2.247×10^{-9}	9.490×10^{-9}
Maxwell	1.500×10^{-9}	–	–
NS	1.897×10^{-4}	1.842×10^{-4}	2.747×10^{-3}
Task	**PDE residual RMS ↓**	**Div. max ↓**	**Time (s) ↓**
Stokes	1.021×10^{-7}	–	117.0 ± 41.6
Maxwell	6.939×10^{-8}	3.056×10^{-16}	0.2270 ± 0.1283
NS	2.731×10^{-3}	–	0.204 ± 0.062

data, error remains 6.415×10^{-13} at $T = 1000$, with maximum relative energy drift 4.276×10^{-12} (Table 11).

The Well datasets. The learned hybrid achieves best mean VRMSE on all five The Well datasets across all three evaluation windows (Table 2; Figure 5). On active matter, one-step mean VRMSE is 0.01668 and median is 0.01237, with rollout means 0.75143 and 1.39739 over steps 1–20 and 21–60. Gains over Flower range from 25% (TRL 2D one-step) to 46% (TRL 3D one-step) and from 10% to 30% on rollout windows. The 17.34M backbone uses eight shared-weight D_4 passes. WaveLiT size and rollout-finetuned comparisons appear in Table 7; these do not establish an accuracy-per-parameter advantage for SP-FNO.

Protocol B uses one training seed, validation-selected checkpoints, and no data augmentation. The symmetry group is selected from $\{\mathrm{Id}, C_4, D_4\}$ on validation data; ensemble passes reuse the same weights and do not measure training-seed uncertainty. The separate autoregressive protocol gives 2.715 ± 0.7669 VRMSE with $K = 3$ refinements; its aggregation differs from the $4 \to 1$ evaluation (Appendix D).

Table 4: Structure and physics ablations: representation ladder, Hodge projection, inf-sup filter, and structure/capacity sweep.

(A) Stokes representation (seed 2026, F=128)

Representation	$L^2(u)\downarrow$	$L^2(p)\downarrow$	Div. RMS $\downarrow$	Bound. RMS $\downarrow$
Collocation projection	1.590×10^{-5}	1.555×10^{-5}	1.371×10^{-4}	4.727×10^{-6}
Continuum div-free	2.548×10^{-5}	1.895×10^{-5}	7.597×10^{-15}	6.873×10^{-6}
+ boundary conforming	$\mathbf{5.573\times10^{-6}}$	$\mathbf{4.422\times10^{-6}}$	$\mathbf{2.033\times10^{-16}}$	$\mathbf{0}$

(B) Hodge projection (K=3)

	Proj. train + proj. eval	Proj. train, no proj. eval	No proj. train + eval
VRMSE	$\mathbf{2.715}$	4.534	2.752
RMSE	0.08009	0.1035	$\mathbf{0.07797}$
Div. RMS	$\mathbf{2.831\times10^{-6}}$	1.816	0.9869

(C) Mixed-Stokes inf-sup filter

	F=8		F=16	
	Before	After	Before	After
β_h	1.308×10^{-4}	1.688	5.847×10^{-5}	0.5722
$\kappa(S)$	8.553×10^{9}	51.34	4.949×10^{10}	516.7
$L^2(p)$	2.698×10^{4}	1.337	1.009×10^{4}	0.3044
$L^2(u)$	11.50	66.15	65.86	138.0

(D) Active-matter structure/capacity ($4\to1$)

Variant	One-step mean $\downarrow$	One-step median $\downarrow$	T[1:20] median $\downarrow$	T[21:60] median $\downarrow$	RMSE $\downarrow$
D1 FNO-only (full proj.)	0.21218	0.13114	1.04019	2.04634	0.04193
D2 SP-FNO transport ref.	0.19340	0.12275	0.95962	1.56354	0.04065
D3 SP-FNO residual ref.	0.18743	0.10887	1.00000	2.18521	0.03386
D4 No projection (diag.)	0.41542	0.23009	0.99471	1.58303	0.07904
D5 Scratch, eff. batch	0.03071	0.02054	0.35227	1.31164	0.00659
D6 Scratch, val. mean	0.02960	0.02001	$\mathbf{0.32799}$	1.29664	0.00645
D7 Refinement (K=4)	0.02733	0.01845	0.33738	$\mathbf{1.29142}$	0.00590
D8 SP-FNO (D_4 ensemble)	$\mathbf{0.01668}$	$\mathbf{0.01237}$	0.46404	1.35240	$\mathbf{0.00454}$

Table 5: Functional majorant reliability on manufactured Poisson and Maxwell problems. Each check compares the float64 majorant M with the measured energy error $\mathcal{E}$ of the manufactured solution; "$M \geq \mathcal{E}$" records how many settings satisfy it.

Problem	# settings	Effectivity index	Checks $M \geq \mathcal{E}$
Manufactured Poisson (Repin)	29	1.000–4.158×10^{3}	$29/29$
Manufactured Maxwell/TE (no anchor)	17	1.783–14.51	$17/17$

Manufactured PDEs and cost. Without a frequency prior, Stokes and Maxwell reach errors near 10^{-9}, and Navier–Stokes reaches 1.90×10^{-4} (Table 3). We average 5, 3, and 3 seeds, respectively. Stokes uses $F = 1024$ features on a laptop CPU, Maxwell $F = 1024$ on an RTX 4090, and Navier–Stokes $F = 64$ on a laptop CPU. Timings include per-instance assembly and solution, excluding reusable offline setup. Across n right-hand sides, total cost is $T_{\text{setup}} + nT_{\text{online}}$; Stokes setup costs about one online solve. These results do not establish a speed advantage over a one-off classical solve. Appendix D gives the cost breakdown.

Ablations. Continuum divergence-free features reduce Stokes divergence RMS to 7.597×10^{-15}, and the boundary envelope removes the boundary residual (Table 4A). Removing Hodge projection in both training and evaluation leaves divergence RMS 0.9869 in protocol A (panel B).

The inf-sup filter improves pressure stability at the cost of discarded constraint directions (panel C). At $F = 8$, it raises β_h from 1.308×10^{-4} to 1.688 and reduces pressure error from 2.698×10^{4} to 1.337, while the first velocity-component error increases from 11.50 to 66.15. At $F = 16$, filtered pressure error falls to 0.3044, but velocity error and Schur conditioning worsen. These coarse, nonconforming bases isolate the stability–approximation tradeoff; their large velocity, divergence, and boundary errors persist after filtering (Table 9). They differ from the $F = 1024$ accuracy benchmark and the $F = 128$ representation ladder.

The cutoffs 10^{-4}, 10^{-3}, and 10^{-2} retain the same six pressure directions at $F = 8$, while cutoffs through 10^{-5} retain the weak seventh direction (Table 18A). This plateau supports the chosen threshold for this basis, without establishing a universal cutoff.

In protocol B, four shared refinement steps lower mean one-step VRMSE from 0.02960 to 0.02733 (Table 4D). The D_4 ensemble reaches 0.01668, but its rollout medians 0.46404 and 1.35240 exceed the refinement configuration's 0.33738 and 1.29142. The one-step gain therefore does not extend to rollout accuracy.

Majorant validation. We compare each majorant with the measured energy error $\mathcal{E} = \|\mathbf{u} - \mathbf{u}_h\|_{\mathcal{A}}$ of the manufactured solution. All 29 Poisson and 17 unanchored Maxwell settings satisfy the float64 check $M \geq \mathcal{E}$ (Table 5). Effectivity $M/\mathcal{E}$ ranges from 1.000 to 4.158×10^3 for Poisson and 1.783 to 14.51 for Maxwell. Dual-space and split-parameter choices affect tightness. We use analytic Gram assembly without interval or outward-rounded arithmetic, so these numerical evaluations do not certify roundoff (Appendix F).

Table 19 reports the primal, dual-space, split-parameter, and flux-prior sweeps. Split-parameter alternation gives effectivity 4.889, compared with 4.914 for the best fixed setting. The manufactured-flux prior yields effectivity near one, but one of 17 settings falls below the measured error by a relative margin of 2.4×10^{-10}; these checks do not enclose numerical error. The analytic Stokes example in Table 6 illustrates that a valid majorant can remain nonzero even when the velocity approximation is exact. With $\Sigma_h = \nu \nabla \mathbf{u}_h$, the stress mismatch vanishes and the infimum over ε recovers the strong-residual bound. A pressure-carrying reconstruction can cancel the equilibrium residual while retaining a nonzero stress mismatch. Minimizing either residual alone therefore need not minimize the full majorant.

5 Limitations

The frozen basis limits accuracy outside its frequency span; a small PDE residual alone cannot detect missing solution components. The learned field-space hybrid adds expressiveness and requires training. Closed-form Gram assembly and the implemented Poincaré and inf-sup constants cover boxes and manufactured problems; general domains require new constants and quadrature or mesh-based assembly. Frequency anchoring appears only in an ablation, never in the headline manufactured results. The The Well comparisons use unequal model and inference budgets across datasets, and the majorant evaluations provide no interval-arithmetic certificate.

Although the reduced matrices are independent of the constraint-block coupling scale, extreme γ can still cause floating-point cancellation in pressure recovery because its right-hand side contains γ^{-1} and γ^{-2} terms.

6 Conclusion

SP-FNO combines frozen spectral features, constrained coefficient solves, and analytically assembled residual majorants. The experiments establish numerical precision within a representable advection span, small errors on unanchored manufactured PDEs, and improved prediction on five The Well datasets with a projected learned hybrid. Constraint preservation depends on the representation; filtering trades pressure resolution for stability, and numerical error certification requires validated constants and arithmetic. The learned hybrid extends capacity with training and inference costs that remain separate from the frozen solver.

AI USE STATEMENT

We used generative AI tools to polish the manuscript's wording, adjust formatting, and assist in reviewing code for correctness and rigor. The authors reviewed all AI-assisted changes and take responsibility for the final manuscript and code.

ETHICS STATEMENT

We study numerical PDE solvers using public simulation data and manufactured benchmarks, with no human participants, personal data, or physical experiments. We state the assumptions and limitations of the error bounds. The authors report no conflicts of interest.

REPRODUCIBILITY STATEMENT

Section 3 specifies the model and coefficient solver. Appendices A, B, and E provide the derivations and proofs; Appendices D and C describe the experimental protocols and active-matter preprocessing. The anonymous repository provides code, environment specifications, configurations, seeds, data manifests, and commands for the experiments: `https://anonymous.4open.science/r/SP-FNO-C44A`. The result records distinguish reproduced experiments from published baseline values.

REFERENCES

Michele Benzi, Gene H. Golub, and Jörg Liesen. Numerical Solution of Saddle Point Problems. *Acta Numerica*, 14:1–137, 2005. doi: 10.1017/S0962492904000212. URL https://doi.org/10.1017/S0962492904000212.

Pavel B. Bochev and Richard B. Lehoucq. On the Finite Element Solution of the Pure Neumann Problem. *SIAM Review*, 47(1):50–66, 2005. doi: 10.1137/S0036144503426074. URL https://doi.org/10.1137/S0036144503426074.

Daniele Boffi, Franco Brezzi, and Michel Fortin. *Mixed Finite Element Methods and Applications*, volume 44 of *Springer Series in Computational Mathematics*. Springer, Berlin/Heidelberg, 2013. doi: 10.1007/978-3-642-36519-5. URL https://doi.org/10.1007/978-3-642-36519-5.

Jingrun Chen, Xurong Chi, Weinan E, and Zhouwang Yang. Bridging Traditional and Machine Learning-Based Algorithms for Solving PDEs: The Random Feature Method. *Journal of Machine Learning*, 1(3):268–298, 2022. doi: 10.4208/jml.220726. URL https://doi.org/10.4208/jml.220726.

Bernardo Cockburn, Guido Kanschat, and Dominik Schötzau. A Note on Discontinuous Galerkin Divergence-free Solutions of the Navier–Stokes Equations. *Journal of Scientific Computing*, 31 (1–2):61–73, 2007. doi: 10.1007/s10915-006-9107-7. URL https://doi.org/10.1007/s10915-006-9107-7.

Chinmay Datar, Taniya Kapoor, Abhishek Chandra, Qing Sun, Erik Lien Bolager, Iryna Burak, Anna Veselovska, Massimo Fornasier, and Felix Dietrich. Fast training of accurate physics-informed neural networks without gradient descent. In *International Conference on Learning Representations*, volume 2026, pp. 104770–104829, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/ab435f7130455143fd9489f389777552-Abstract-Conference.html.

Suchuan Dong and Zongwei Li. Local Extreme Learning Machines and Domain Decomposition for Solving Linear and Nonlinear Partial Differential Equations. *Computer Methods in Applied Mechanics and Engineering*, 387:114129, 2021. doi: 10.1016/j.cma.2021.114129. URL https://doi.org/10.1016/j.cma.2021.114129.

Howard C. Elman, David J. Silvester, and Andrew J. Wathen. *Finite Elements and Fast Iterative Solvers: with Applications in Incompressible Fluid Dynamics*. Oxford University Press, Oxford, 2005. doi: 10.1093/oso/9780198528678.001.0001. URL https://academic.oup.com/book/54449.

Lewin Ernst, Nikolaos Rekatsinas, and Karsten Urban. A posteriori certification of PDE approximations with particular application to neural networks. *arXiv preprint arXiv:2502.20336*, 2025. URL https://arxiv.org/abs/2502.20336.

Vladimir Fanaskov, Tianchi Yu, Alexander Rudikov, and Ivan Oseledets. Astral: Training Physics-Informed Neural Networks with Error Majorants. *arXiv preprint arXiv:2406.02645*, 2024. URL https://arxiv.org/abs/2406.02645. Accepted at the ICLR 2026 Workshop on AI and Partial Differential Equations.

Juncai He, Xinliang Liu, and Zitong Tian. Divergence-free Linearized Neural Networks: Integral Representation and Optimal Approximation Rates. *arXiv preprint arXiv:2603.28638*, 2026. URL https://arxiv.org/abs/2603.28638.

Qifeng Hu and Inanc Senocak. Unsupervised simulation of incompressible flows with physics- and equality- constrained artificial neural networks. *arXiv preprint arXiv:2511.18820*, 2025. URL https://arxiv.org/abs/2511.18820.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural Operator: Learning Maps Between Function Spaces With Applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL https://jmlr.org/papers/v24/21-1524.html.

Aditi S. Krishnapriyan, Amir Gholami, Shandian Zhe, Robert M. 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. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/hash/df438e5206f31600e6ae4af72f2725f1-Abstract.html.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier Neural Operator for Parametric Partial Differential Equations. In *International Conference on Learning Representations (ICLR)*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Xiaotian Liu, Shuyuan Shang, Xiaopeng Wang, Pu Ren, and Yaoqing Yang. Iterative Refinement Neural Operators are Learned Fixed-Point Solvers: A Principled Approach to Spectral Bias Mitigation. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2605.24041.

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. doi: 10.1038/s42256-021-00302-5. URL https://www.nature.com/articles/s42256-021-00302-5.

Till Muser, Alexandra Spitzer, Matti Lassas, Maarten V. de Hoop, and Ivan Dokmanić. Flowers: A Warp Drive for Neural PDE Solvers. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2603.04430.

Tung Nguyen, Arsh Koneru, Shufan Li, and Aditya Grover. PhysiX: A Foundation Model for Physics Simulations. *arXiv preprint arXiv:2506.17774*, 2025. doi: 10.48550/arXiv.2506.17774. URL https://arxiv.org/abs/2506.17774.

Ruben Ohana, Michael McCabe, Lucas Meyer, Rudy Morel, Fruzsina J. Agocs, Miguel Beneitez, Marsha Berger, Blakesley Burkhart, Stuart B. Dalziel, Drummond B. Fielding, Daniel Fortunato, Jared A. Goldberg, Keiya Hirashima, Yan-Fei Jiang, Rich R. Kerswell, Suryanarayana Maddu, Jonah Miller, Payel Mukhopadhyay, Stefan S. Nixon, Jeff Shen, Romain Watteaux, Bruno Régaldo-Saint Blancard, François Rozet, Liam H. Parker, Miles Cranmer, and Shirley Ho. The Well: a Large-Scale Collection of Diverse Physics Simulations for Machine Learning. In *Advances in Neural Information Processing Systems*,

volume 37, pp. 44989–45037. Curran Associates, Inc., 2024. doi: 10.52202/079017-1430. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/4f9a5acd91ac76569f2fe291b1f4772b-Abstract-Datasets_and_Benchmarks_Track.html.

Youngjae Park, Jaemin Kim, and Junghwa Hong. Coupling-Robust Accuracy in Multiphysics Physics Informed Neural Networks via Kronecker-Preconditioned Optimization. *arXiv preprint arXiv:2605.23391*, 2026. URL https://arxiv.org/abs/2605.23391. Extended version of a paper accepted at the ICML 2026 AI4Physics Workshop.

Ali Rahimi and Benjamin Recht. Random Features for Large-Scale Kernel Machines. In *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007. URL https://papers.nips.cc/paper/2007/hash/013a006f03dbc5392effeb8f18fda755-Abstract.html.

Maziar Raissi, Paris Perdikaris, and George Em 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. doi: 10.1016/j.jcp.2018.10.045. URL https://doi.org/10.1016/j.jcp.2018.10.045.

Fei Ren, Sifan Wang, Pei-Zhi Zhuang, Hai-Sui Yu, and He Yang. General Fourier Feature Physics-Informed Extreme Learning Machine (GFF-PIELM) for High-Frequency PDEs. *arXiv preprint arXiv:2510.12293*, 2025. URL https://arxiv.org/abs/2510.12293.

Sergey I. Repin. *A Posteriori Estimates for Partial Differential Equations*, volume 4 of *Radon Series on Computational and Applied Mathematics*. De Gruyter, Berlin, New York, 2008. doi: 10.1515/9783110203042. URL https://www.degruyterbrill.com/document/doi/10.1515/9783110203042/html.

Pratanu Roy and Stephen T. Castonguay. Exact enforcement of temporal continuity in sequential physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 430:117197, 2024. doi: 10.1016/j.cma.2024.117197. URL https://doi.org/10.1016/j.cma.2024.117197.

Shyam Sankaran, Hanwen Wang, and Paris Perdikaris. Small Models, Strong Priors: Architectural Inductive Bias for Parameter-Efficient Neural PDE Solvers. *arXiv preprint arXiv:2605.25949*, 2026. URL https://arxiv.org/abs/2605.25949.

Antonin Sulc. FastLSQ: Solving PDEs in One Shot via Fourier Features with Exact Analytical Derivatives. *arXiv preprint arXiv:2602.10541*, 2026. URL https://arxiv.org/abs/2602.10541. Accepted at the ICLR 2026 Workshop on AI and Partial Differential Equations.

Sifan Wang, Shyam Sankaran, and Paris Perdikaris. Respecting Causality for Training Physics-Informed Neural Networks. *Computer Methods in Applied Mechanics and Engineering*, 421:116813, 2024. doi: 10.1016/j.cma.2024.116813. URL https://doi.org/10.1016/j.cma.2024.116813.

Xiong Xiong, Ruonan Zhai, Zheng Zeng, Sheng Zhou, Rongchun Hu, and Zichen Deng. Frequency Shift Physics-Informed Extreme Learning Machine for Solving High-Frequency Partial Differential Equations. *arXiv preprint arXiv:2607.01694*, 2026. URL https://arxiv.org/abs/2607.01694.

A CONSTRAINT ELIMINATION AND MIXED-SYSTEM SOLVER

A.1 PROJECTOR AND NUMERICAL RANK

For the exact rank-r SVD $D = U_r \Sigma_r V_r^\top$, $D^\dagger D = V_r V_r^\top$ projects onto $\mathrm{range}(D^\top)$. Its orthogonal complement is $\ker(D)$, proving Proposition 1. The identity $D^\dagger = D^\top (DD^\top)^{-1}$ requires full row rank; the SVD formula also covers overdetermined and rank-deficient constraints.

Equation equation 4 quantifies the numerical constraint residual when singular values are discarded. Exact continuum feasibility additionally requires a conforming representation.

A.2 INHOMOGENEOUS CONSTRAINTS AND PRESSURE RECOVERY

After pressure gauging and filtering, define $B_r = T_p^\top B$ and $S = B_r A^{-1} B_r^\top$. For $\gamma B_r \mathbf{c_v} = \mathbf{g}_r$, a particular solution is

$$\mathbf{c}_{\mathrm{part}} = A^{-1} B_r^\top S^{-1} (\gamma^{-1} \mathbf{g}_r).$$

The remaining component lies in $\ker(B_r)$ and solves equation 7. Eliminating velocity from the first row of equation 6 gives equation 8. Algorithm 2 uses linear solves for the displayed inverses.

Algorithm 2 Projected velocity solve and Schur pressure recovery

Require: $A \succ 0$, pressure mass M_p, coupling B, loads $\mathbf{f}_v, \mathbf{g}_p$, $\gamma > 0$, cutoff τ

Form an orthonormal zero-mean pressure basis G_p
$B_0 \leftarrow G_p^\top B$; $M_0 \leftarrow G_p^\top M_p G_p$
$L_v L_v^\top \leftarrow A$; $L_p L_p^\top \leftarrow M_0$
$U \Sigma V^\top \leftarrow \mathrm{SVD}(L_p^{-1} B_0 L_v^{-\top})$
Retain $\{i : \sigma_i > \max(\tau \sigma_1, \tau_{\mathrm{abs}})\}$
$T_p \leftarrow L_p^{-\top} U_r$; $B_r \leftarrow T_p^\top B_0$
$\mathbf{g}_r \leftarrow T_p^\top G_p^\top \mathbf{g}_p$
Verify that B_r has full row rank; form $S \leftarrow B_r A^{-1} B_r^\top$
$Z \leftarrow$ orthonormal null-space basis of B_r
$\mathbf{c}_{\mathrm{part}} \leftarrow A^{-1} B_r^\top S^{-1} (\gamma^{-1} \mathbf{g}_r)$
Solve $(Z^\top A Z) \mathbf{c}_z = Z^\top (\mathbf{f}_v - A \mathbf{c}_{\mathrm{part}})$
Solve $S \mathbf{p}_r = \gamma^{-1} B_r A^{-1} \mathbf{f}_v - \gamma^{-2} \mathbf{g}_r$
return $\mathbf{c_v} = \mathbf{c}_{\mathrm{part}} + Z \mathbf{c}_z$, $\mathbf{c}_p = G_p T_p \mathbf{p}_r$

A.3 PROOF OF DECOUPLED CONDITIONING

For orthonormal Z, Rayleigh–Ritz bounds give

$$\lambda_{\min}(A) \|\mathbf{z}\|^2 \le \mathbf{z}^\top Z^\top A Z \mathbf{z} \le \lambda_{\max}(A) \|\mathbf{z}\|^2 .$$

Hence $Z^\top A Z \succ 0$ and $\kappa_2(Z^\top A Z) \le \kappa_2(A)$. From $B_r = U_r^\top \widehat{B} L_v^\top$ and $A = L_v L_v^\top$,

$$S = U_r^\top \widehat{B} \widehat{B}^\top U_r = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_r^2).$$

The retained singular values are positive, so $S \succ 0$ and $\kappa_2(S) = (\sigma_1/\sigma_r)^2 \le \tau^{-2}$. Neither matrix contains γ, proving Proposition 2. The right-hand side of equation 8 contains γ^{-1} and γ^{-2}, so extreme scaling can cause floating-point cancellation despite fixed conditioning.

The coupling experiment in Table 18(B) uses 12 velocity and five pressure dimensions, duplicated coupling rows, and $\tau = 10^{-3}$, retaining three pressure directions. Across $\gamma \in \{1, 10, 100, 1000, 10000\}$, $\kappa_2(Z^\top A Z) = 1.805$, while the monolithic condition number increases from 8.130 to 1.354×10^4 and the symmetric square-root block-scaled condition number increases to 136.2.

Result 1 (Mixed-Stokes filtering experiment)**.** *For seed 2026 and $F = 8$ velocity and pressure features per component, including a pressure mode at $\omega = (500, 550)$, filtering at $\tau = 10^{-4}$ retains six of seven gauged pressure directions. The discrete inf-sup constant increases from 1.308×10^{-4} to 1.688, the Schur condition number decreases from 8.553×10^9 to 51.34, and measured relative pressure error falls from 2.698×10^4 to 1.337. Filtering improves stability, but the pressure error remains substantial.*

Table 6: Analytic Stokes majorants for the polynomial example. $\mathcal{E} = \|\mathbf{u} - \mathbf{u}_h\|_{\mathcal{A}}$ is the energy error; M_{sr} and M_{fs} bound the energy norm, not its square. Values are rounded from exact integrals.

Trial	$\mathcal{E}$	M_{sr}	M_{fs}	$M_{\mathrm{fs}}/M_{\mathrm{sr}}$
$\mathbf{u}_h = \mathbf{u}$, $\Sigma_h = \nu\nabla\mathbf{u}_h$	0	0.3898	0.3898	1.000
$\mathbf{u}_h = \mathbf{u}$, $\Sigma_h = \nu\nabla\mathbf{u}_h - pI$	0	0.3898	0.7071	1.814
$\mathbf{u}_h = \frac{3}{4}\mathbf{u}$, $\Sigma_h = \nu\nabla\mathbf{u}_h$	0.00825	0.3902	0.3902	1.000

B Functional majorant derivation

B.1 Stokes velocity estimate

Consider $-\nu\Delta\mathbf{u} + \nabla p = \mathbf{f}$ with homogeneous Dirichlet velocity and $\nabla \cdot \mathbf{u} = 0$. Under (D0) of Proposition 3, $e = \mathbf{u} - \mathbf{u}_h$ belongs to $H_0^1(\Omega)^d$ and is divergence-free. Testing the weak form with e eliminates pressure. For $\Sigma_h \in H(\mathrm{Div}; \Omega)$, integration by parts gives

$$\begin{aligned}\nu \|\nabla e\|_{L^2}^2 &= \int_\Omega \mathbf{f} \cdot e - \nu \int_\Omega \nabla\mathbf{u}_h : \nabla e \\ &= \int_\Omega (\nabla \cdot \Sigma_h + \mathbf{f}) \cdot e + \int_\Omega (\Sigma_h - \nu\nabla\mathbf{u}_h) : \nabla e.\end{aligned}$$

The zero trace of e removes the boundary pairing without restricting $\Sigma_h n$. Cauchy–Schwarz and $\|e\|_{L^2} \le C_\Omega \|\nabla e\|_{L^2}$ yield

$$\|e\|_{\mathcal{A}}^2 \le \nu^{-1} \left(\|\Sigma_h - \nu\nabla\mathbf{u}_h\|_{L^2} + C_\Omega \|\nabla \cdot \Sigma_h + \mathbf{f}\|_{L^2}\right)^2.$$

Applying $(a+b)^2 \le (1+\varepsilon)a^2 + (1+\varepsilon^{-1})b^2$ proves equation 11. The factor ν^{-1} follows from the energy norm. The proof also covers nonzero boundary data when exact and approximate velocities share a trace, as with a common lifting.

The exact total stress is $\tau = \nu\nabla\mathbf{u} - pI$ and satisfies $\nabla \cdot \tau + \mathbf{f} = 0$. At $\Sigma_h = \tau$, the mismatch equals $\nu\nabla e - pI$, so the full-stress majorant can remain positive even for exact velocity. Thus the bound need not be sharp. The equilibrium residual in equation 11 is $\nabla \cdot \Sigma_h + \mathbf{f}$: diffusion enters through the reconstructed stress. The Stokes experiments instead evaluate the strong-residual estimate described in Appendix F.

B.2 Analytic comparison of two Stokes majorants

On $\Omega = [0,1]^2$, take $\nu = 1/3$, $\psi = x^2(1-x)^2y^2(1-y)^2$, $\mathbf{u} = (\partial_y\psi, -\partial_x\psi)$, $p = x - 1/2$, and $p_h = 0$. The polynomial gives zero velocity trace and continuous divergence-freedom; set $\mathbf{f} = -\nu\Delta\mathbf{u} + \nabla p$ and $C_\Omega = 1/(\pi\sqrt{2})$. Table 6 compares $M_{\mathrm{sr}} = C_\Omega \nu^{-1/2} \|\mathbf{f} + \nu\Delta\mathbf{u}_h\|_{L^2}$ with the energy-norm form of equation 11:

$$M_{\mathrm{fs}} = \sqrt{\inf_{\varepsilon>0} \mathcal{M}_\varepsilon} = \nu^{-1/2} \left(\|\Sigma_h - \nu\nabla\mathbf{u}_h\|_{L^2} + C_\Omega \|\nabla \cdot \Sigma_h + \mathbf{f}\|_{L^2}\right).$$

For $\Sigma_h = \nu\nabla\mathbf{u}_h$, the mismatch vanishes and $M_{\mathrm{fs}} = M_{\mathrm{sr}}$. At exact velocity, the total stress $\Sigma_h = \nu\nabla\mathbf{u} - pI$ instead gives zero equilibrium residual and mismatch $-pI$, yielding $M_{\mathrm{fs}} = 1/\sqrt{2}$ despite zero velocity error. Thus exact stress equilibration need not sharpen this velocity bound.

Polynomial integration gives $\|\nabla\mathbf{u}\|_{L^2}^2 = 4/1225$, $\|\Delta\mathbf{u}\|_{L^2}^2 = 128/525$, and $(\Delta\mathbf{u}, \nabla p)_{L^2} = 0$. For $\mathbf{u}_h = \frac{3}{4}\mathbf{u}$, these give $\mathcal{E} = \sqrt{3}/210$ and $M_{\mathrm{sr}} = \sqrt{66262}/(210\pi)$. These are analytic illustrations, separate from the numerical experiments.

B.3 Coupled quadratic assembly

Write $\Sigma_h(x) = \Sigma(x)\mathbf{s}$. The tensor-valued features may include pressure-type modes qI. With analytic L^2 Gram weighting, write the mismatch and equilibrium residuals as $A_1\mathbf{s} - a_1$ and $A_2\mathbf{s} - a_2$,

where a_1 represents $\nu\nabla\mathbf{u}_h$ and a_2 represents $-\mathbf{f}$. Then

$$B_\varepsilon = \nu^{-1}\left[(1+\varepsilon)A_1^\top A_1 + (1+\varepsilon^{-1})C_\Omega^2 A_2^\top A_2\right],$$
$$\mathbf{b}_\varepsilon = \nu^{-1}\left[(1+\varepsilon)A_1^\top a_1 + (1+\varepsilon^{-1})C_\Omega^2 A_2^\top a_2\right],$$
$$c_\varepsilon = \nu^{-1}\left[(1+\varepsilon)\|a_1\|^2 + (1+\varepsilon^{-1})C_\Omega^2\|a_2\|^2\right].$$

A single solve in $\mathbf{s}$ minimizes both residual terms. If the dual features are linearly independent in L^2, $B_\varepsilon \succ 0$ and

$$\mathbf{s}^\star = B_\varepsilon^{-1}\mathbf{b}_\varepsilon, \qquad \mathcal{M}_\varepsilon^\star = c_\varepsilon - \mathbf{b}_\varepsilon^\top B_\varepsilon^{-1}\mathbf{b}_\varepsilon.$$

Each $\varepsilon > 0$ gives an upper bound. The infimum $\mathcal{M}^\star = \inf_{\varepsilon>0}\mathcal{M}_\varepsilon^\star$ also bounds the squared energy error, whether or not the infimum is attained.

B.4 Pressure and numerical validity

To obtain equation 12, let $q = p - p_h$ and $e = \mathbf{u} - \mathbf{u}_h$. Subtracting the approximate momentum equation from the weak form gives, for $\mathbf{v} \in H_0^1(\Omega)^d$,

$$(q, \nabla\cdot\mathbf{v}) = \nu(\nabla e, \nabla\mathbf{v}) - (r_h, \mathbf{v}).$$

Cauchy–Schwarz and Poincaré bound the right-hand side by $(\nu\|\nabla e\|_{L^2} + C_\Omega\|r_h\|_{L^2})\|\nabla\mathbf{v}\|_{L^2}$. Taking the supremum over $\mathbf{v}$ and applying the continuous inf-sup inequality proves the pressure estimate. Both pressures must share the zero-mean gauge. Pressure estimation requires a proven continuous β in the $\|\nabla\mathbf{v}\|_{L^2}$ test norm. The discrete filtering spectrum and dual-stress equilibrium residual do not establish this bound.

Analytic Gram assembly removes quadrature error for the supported features, forcing, and domains. The continuum bounds require the stated admissibility and proven constants. Our float64 evaluations use no interval or outward-rounded arithmetic and do not certify roundoff.

C Active-matter hybrid and symmetry action

The active-matter state contains the 11 physical channels $(c, u_x, u_y, D_{xx}, D_{xy}, D_{yx}, D_{yy}, E_{xx}, E_{xy}, E_{yx}, E_{yy})$. The operator predicts one state from four input states. Let $N(z) = (z-\mu)/\sigma$ denote channel normalization and $\rho(g) = N \circ g_{\text{phys}} \circ N^{-1}$ the normalized action of $g \in D_4$. Rotations and reflections act on vector and tensor components in physical coordinates and do not commute with channel-wise affine normalization.

For each nonzero Fourier wave vector on the 10×10 physical box, Π_{phys} applies $I - kk^\top/|k|^2$ to velocity. It symmetrizes D and sets its trace to c, and symmetrizes E and sets its trace to zero. Projection acts in physical units because these trace relations become affine under channel normalization.

The $K = 4$ refinement configuration uses one 17.34M-parameter backbone pass and four passes of a shared 2.17M-parameter refiner. The D_4 ensemble uses eight equal-weight passes of the shared 17.34M backbone. We select from the fixed candidate groups $\{\text{Id}, C_4, D_4\}$ using validation data. Inference budgets differ, and the ensemble's lower one-step error does not imply lower rollout medians (Table 4).

D Experimental protocols and extended comparisons

Advection. We use frequencies $\omega = 1, \ldots, n$, each as a sine/cosine pair with random phases θ and $\theta + \pi/2$. Their span is $\text{span}\{\sin(kx), \cos(kx)\}_{k=1}^n$. Since $\sin(x - \beta t) = \cos(\beta t)\sin x - \sin(\beta t)\cos x$, the exact solution lies in this span at all times. The main benchmark uses seven pairs (14 functions); the out-of-distribution cases in Table 23 use 16 pairs. The main absolute error stays below 4×10^{-13}. For $\sin(9x) + 0.5\sin(17x)$ the omitted 17th harmonic gives relative error $\sqrt{0.2}$ independently of β. Spectral truncation also limits the broad-spectrum wave packet to 3.96×10^{-3} error.

Table 7: Complete WaveLiT configuration comparison on active matter under the $4 \rightarrow 1$ protocol. All entries are median VRMSE; WaveLiT values follow Sankaran et al. (2026). WaveLiT uses one backbone pass, whereas SP-FNO uses eight shared-weight passes of a 17.34M backbone.

Method	One-step ↓	T[1:20] ↓	T[21:60] ↓
WaveLiT-1.2M-PT	0.0211	0.5098	1.4186
WaveLiT-1.2M-FT	0.0282	0.4072	1.3863
WaveLiT-9.5M-PT	**0.0114**	**0.2096**	**1.3016**
WaveLiT-9.5M-FT	0.0131	0.2231	1.3204
SP-FNO (ours)	0.01237	0.46404	1.35240

Table 8: Published active-matter one-step VRMSE and iterative refinement. PhysiX uses Nguyen et al. (2025); IRNO results use Liu et al. (2026); WaveLiT and WNO use Sankaran et al. (2026). WaveLiT entries are medians, while SP-FNO entries are means; other entries retain their source reporting. Different histories, aggregation rules, capacities, and inference budgets preclude a uniform ranking.

Method	One-step VRMSE ↓
PhysiX (2025)	0.0904
WNO (2022)	0.5736
HFS-ResUNet + IRNO	0.0486
FNO + IRNO (2026)	0.0501
TFNO + IRNO (2026)	0.0387
WaveLiT-1.2M-PT (2026)	0.0211
WaveLiT-1.2M-FT (2026)	0.0282
SP-FNO (ours)	**0.01668**

Method	VRMSE ↓
FNO (2021)	0.1017
FNO + IRNO (2026)	0.0501 (K=6)
TFNO (2025)	0.1981
TFNO + IRNO (2026)	0.0387 (K=4)
SP-FNO + IRNO (ours)	**0.02733** (K=4)

The published advection baselines in Table 1 follow Table 1 of Datar et al. (2026). Their PINN uses four hidden layers and 5,000 epochs, with learning rate 0.1 and batch size 200; the source protocol lists widths $\{10, 20, 30, 40\}$. The separate multi-metric comparison in Table 20 follows that work's Table 7: the L-BFGS model at learning rate 0.1 has ten hidden layers, width 30, and 20,000 epochs. The two PINN comparisons therefore use different training budgets.

The Well protocols and aggregation. Protocol A (Table 4B and Tables 17, 25) uses the public FNO-active-matter backbone (Ohana et al., 2024) with a learned refiner. Each autoregressive frame uses K refinement iterations, evaluated over paired seeds. At $K = 3$, VRMSE is 2.715 ± 0.7669. Protocol B uses The Well $4 \rightarrow 1$ setting and yields one-step VRMSE 0.01668 on active matter (Table 2; Table 4D). All entries in Table 2 are mean VRMSE following Muser et al. (2026). Panel D gives test-sample means and medians for one-step predictions and trajectory medians for rollouts. The corresponding active-matter rollout means are 0.75143 and 1.39739. We use no data augmentation. Protocols A and B have different evaluation settings and are not directly comparable. Table 8 collects published one-step and refinement results; capacity, history length, and inference budget differ across methods.

Table 2 configurations. The mean baselines use one backbone pass and approximately 18.93M (FNO), 18.57M (CNextU-Net), 17.78M (scOT), and 17.33M (Flower) parameters, following Muser et al. (2026). All Table 2 rows use four input frames. SP-FNO uses a 17.34M backbone with eight shared-weight passes. Table 7 retains the complete WaveLiT size and training-stage comparison for active matter. PT denotes the one-step-trained checkpoint; FT adds autoregressive rollout finetuning. The compact PT configuration serves as the reference because its training stage matches the one-step prediction task. WaveLiT-9.5M-PT has lower median errors in all three windows than SP-FNO, and the compact FT variant has lower short-rollout error. Thus the WaveLiT comparison does not imply superiority over the full WaveLiT family or equal-budget superiority.

Representation and Hodge ablations. Panel A of Table 4 holds features, geometry, and solver fixed while comparing Stokes representations: collocation projection, continuum divergence-free stream functions, and stream functions with a boundary-conforming envelope. These correspond to

Table 9: Full mixed-Stokes filtering diagnostics (seed 2026) on coarse random Fourier bases. The algebraic divergence residual is distinct from the continuous divergence and boundary norms.

	F=8		F=16	
	Before	After	Before	After
$L^2(u)$	11.50	66.15	65.86	138.0
$L^2(v)$	14.45	9.74	11.45	9.02
$L^2(p)$	2.698×10^4	1.337	1.009×10^4	0.3044
$\|\nabla \cdot u_h\|$	79.57	79.68	81.00	81.01
$\|u_h\|_{\partial\Omega}$	60.93	188.2	188.1	384.7
Div. resid.	7.7×10^{-14}	8.4×10^{-14}	5.9×10^{-14}	1.7×10^{-13}
$\beta_h \uparrow$	1.308×10^{-4}	**1.688**	5.847×10^{-5}	**0.5722**
Schur cond. $\downarrow$	8.553×10^9	51.34	4.949×10^{10}	516.7
SPD cond.	4443	8353	6.383×10^5	8.188×10^5

a collocation least-squares treatment of the divergence constraint, a null-space construction (Benzi et al., 2005; Boffi et al., 2013), and a divergence-free spectral basis with exact non-penetration (Cockburn et al., 2007). Panel B tests Hodge projection at $K = 3$ under protocol A. Projection in training and evaluation gives VRMSE 2.715; removing projection at evaluation gives 4.534. Removing it in both phases gives 2.752 and divergence RMS 0.9869. Panel C reports the mixed-Stokes inf-sup filter of Section 3.3 on coarse random bases at F=8 and F=16, showing the stability–approximation tradeoff: the filtered pressure error drops ($1.337 \to 0.3044$) while the retained coupling weakens ($\beta_h = 1.688 \to 0.5722$). Table 9 gives all velocity, pressure, constraint, and conditioning diagnostics.

Capacity and symmetry ablations. Panel D's first block uses the public FNO-active-matter backbone (about 19.0M parameters): no refiner, a Flower transport refiner with 17.34M trainable parameters, or an IRNO residual refiner with 0.14M trainable parameters. We freeze the backbone for both learned refiners. The D4 row disables projection in the same backbone family. In the second block, we train the 17.34M-parameter Flowers architecture (Muser et al., 2026) from scratch on the full train split, using effective batch size 112 or validation-mean selection. We then freeze the backbone and apply $K = 4$ shared refinement steps with a 2.17M-parameter refiner. We select the final symmetry group on validation data from $\{\mathrm{Id}, C_4, D_4\}$. Its eight equal-weight backbone passes differ in cost from D7's one backbone pass and four refiner passes.

Manufactured PDE protocols. Table 3 averages five Stokes seeds at $F = 1024$ on a laptop CPU, three Maxwell seeds at $F = 1024$ on an RTX 4090, and three Navier–Stokes seeds at $F = 64$ on a laptop CPU. Table 24 gives Maxwell error standard deviations. Timings include analytic assembly and the dense solve for one instance, excluding amortized offline SVD setup. Stokes reaches 2.14×10^{-9} relative L^2 error; Maxwell reaches 1.50×10^{-9} with maximum divergence 3.06×10^{-16}; Navier–Stokes reaches 1.90×10^{-4}.

Compute cost. Offline work comprises frequency sampling, analytic Gram assembly, and the rank-revealing SVD projector and null basis of Proposition 1. The SVD costs $O(m(KN)\min(m, KN))$ and is amortized across instances. Online work comprises a dense least-squares or reduced-SPD solve with KN unknowns, basis evaluation, and, for Navier–Stokes, residual assembly and Newton updates. Steady frozen-feature solves require neither gradient descent nor time stepping.

The dense solves account for 0.092 ± 0.025s of the 0.204 ± 0.062s Navier–Stokes total; basis evaluation and Newton residuals account for the remainder. For Stokes, dense solves account for 33.15 ± 13.40s of 117.0 ± 41.6s, with the remainder spent on basis evaluation, collocation assembly, and coefficient-space projection. Both totals exclude offline setup. At $F = 128$, CUDA solves take 0.0097–0.0194s on an RTX 4090 (Table 14). At the Stokes $F = 1024$ setting, offline setup costs about as much as one instance solve on a laptop CPU; these timings do not establish an advantage over a one-off direct classical solve.

Majorant validation. Table 5 reports the ratio of each majorant to the true energy error. All 29 Poisson and 17 Maxwell settings satisfy the upper-bound check. The sweeps vary primal and dual spaces and the split parameter, with Poisson effectivities up to 4.158×10^3. For Stokes, we report the strong-residual and Poincaré bound $\|e\|_{\mathcal{A}} \le C_\Omega \nu^{-1/2} \|\mathbf{f} - \mathcal{L}\mathbf{u}_h\|_{L^2}$ of Appendix F. Equation equation 11 gives the dual-field functional bound realized by analytic Gram assembly. The inequalities hold for the analytic quantities; float64 evaluation without interval or outward-rounded arithmetic does not certify roundoff.

E INF-SUP FILTERING

This section derives the filtering guarantee and specifies the settings for Result 1.

Setup. Let $\{\phi_j\}_{j=1}^{N_v}$ and $\{\psi_a\}_{a=1}^{N_p}$ be the frozen velocity and pressure Fourier bases on $\Omega = [0,1]^2$, and let

$$A_{ij} = \nu \int_\Omega \nabla\phi_i : \nabla\phi_j \, dx, \qquad (M_p)_{ab} = \int_\Omega \psi_a \psi_b \, dx, \qquad B_{aj} = -\int_\Omega \psi_a \, \nabla\cdot\phi_j \, dx.$$

We assemble these matrices analytically (Section F) and remove the constant pressure mode. With an SVD basis G_p of $\ker(\mathbf{1}_p^\top M_p)$, the zero-mean gauge gives $B \leftarrow G_p^\top B$ and $M_p \leftarrow G_p^\top M_p G_p$, removing one scalar direction.

Discrete inf-sup constant. Let $A = L_v L_v^\top$ and $M_p = L_p L_p^\top$ be Cholesky factorizations. The whitened coupling is

$$\widehat{B} = L_p^{-1} B L_v^{-\top}. \tag{14}$$

Its smallest singular value, including zeros for uncoupled pressure directions, is the discrete inf-sup constant:

$$\beta_h = \inf_{\mathbf{q}\neq 0} \sup_{\mathbf{v}\neq 0} \frac{\mathbf{q}^\top B \mathbf{v}}{\|\mathbf{v}\|_A \, \|\mathbf{q}\|_{M_p}} = \sigma_{\min}(\widehat{B}). \tag{15}$$

The substitutions $\mathbf{v} = L_v^{-\top}\mathbf{z}_v$ and $\mathbf{q} = L_p^{-\top}\mathbf{z}_p$ reduce the quotient to Euclidean norms. The pressure-mass-whitened Schur complement satisfies

$$L_p^{-1} S L_p^{-\top} = \widehat{B}\widehat{B}^\top, \qquad S = BA^{-1}B^\top,$$

so its smallest eigenvalue is β_h^2. A small positive β_h indicates weak discrete stability; $\beta_h = 0$ makes the saddle system singular.

Whitened filtering. Let $U\Sigma V^\top = \widehat{B}$. We retain the r pressure directions with $\sigma_i > \max(\tau\sigma_{\max}, \tau_{\text{abs}})$, using $\tau = 10^{-4}$ and $\tau_{\text{abs}} = 0$ in the mixed-Stokes experiment. The pressure transformation is

$$T_p = L_p^{-\top} U_{:r}, \qquad \mathbf{q} = T_p \mathbf{q}_r, \tag{16}$$

with $T_p^\top M_p T_p = I_r$. The retained coupling $B_r = T_p^\top B$ satisfies $B_r L_v^{-\top} = U_{:r}^\top \widehat{B}$, retaining singular values $\sigma_1, \dots, \sigma_r$ in the whitened norms. For $\gamma \neq 0$, the rotated saddle system reads

$$\begin{pmatrix} A & \gamma B_r^\top \\ \gamma B_r & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{q}_r \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix}, \tag{17}$$

Its unscaled Schur complement is

$$S_r = B_r A^{-1} B_r^\top = U_{:r}^\top \widehat{B}\widehat{B}^\top U_{:r} = \operatorname{diag}(\sigma_1^2, \dots, \sigma_r^2).$$

Proposition 4 (Stability recovery)**.** *Assume $A, M_p \succ 0$, $\gamma \neq 0$, and retain $r \ge 1$ positive singular values $\sigma_1 \ge \cdots \ge \sigma_r > \tau\sigma_{\max}$ of the whitened coupling after the zero-mean gauge. Then the filtered pair has $\beta_h = \sigma_r > 0$, its saddle system has a unique solution, and $\kappa(S_r) = (\sigma_1/\sigma_r)^2$.*

Proof. The filtered inf-sup constant is

$$\sigma_{\min}(T_p^\top B L_v^{-\top}) = \sigma_{\min}(U_{:r}^\top \widehat{B}) = \sigma_r.$$

Thus B_r has full row rank. Together with $A \succ 0$, this gives the Brezzi conditions and a nonsingular saddle matrix. The factorization of S_r gives the spectrum $\{\sigma_1^2, \dots, \sigma_r^2\}$ and the condition number. □

For a conforming Galerkin approximation of the filtered variational problem, the mixed approximation estimate of Elman et al. (2005) controls the error by the best approximation errors in the velocity energy and pressure-mass norms, with constants depending on coercivity, continuity, and β_h^{-1}. The pressure-mass norm becomes Euclidean in the coordinates $\mathbf{q}_r$ because $T_p^\top M_p T_p = I_r$. This stability result does not bound the approximation error from discarding pressure directions.

Mixed-Stokes experiment. We use seed 2026, $F = 8$ features per velocity component, 8 pressure features, and one high-frequency pressure mode at $\omega \approx (500, 550)$. We evaluate on 5000 random points in float64:

- Unfiltered ($\tau = 0$): 7 pressure directions remain after the gauge, $\beta_h^{\mathrm{raw}} = 1.308 \times 10^{-4}$, $\kappa(S) = 8.553 \times 10^9$, and relative L^2 pressure error 2.698×10^4.
- Filtered ($\tau = 10^{-4}$): 6 directions remain, $\beta_h = 1.688$, $\kappa(S_r) = 51.34$, and relative L^2 pressure error 1.337 (Table 18, panel A).

The rounded singular values are $\sigma_1 = 12.09$, $\sigma_7 = 1.308 \times 10^{-4}$, and $\sigma_6 = 1.688$, giving $\lambda_{\min}(S_r) = 2.849$. The unrounded values give $\kappa(S_r) = 51.34$; the rounded ratio gives $(12.09/1.688)^2 \approx 51.3$. Table 18 retains the same dimension for $\tau \in [10^{-4}, 10^{-2}]$. The default $\tau = 10^{-8}$ retains all 7 directions. We use 10^{-4}, the first tested threshold that removes the unstable direction.

Spectral filter for collocation. For the collocation-projection Stokes/NS solver, we parameterize pressure independently and enforce divergence at collocation points. We use a heuristic spectral filter with default threshold $\tau = 10^{-6}$: compute an SVD of the velocity-basis divergence matrix D, form $P_{\mathrm{range}(D^\top)} = V_r V_r^\top$, and drop pressure basis functions whose gradient column G_j satisfies $\left\|P_{\mathrm{range}(D^\top)} G_j\right\| < \tau \max_j \left\|P_{\mathrm{range}(D^\top)} G_j\right\|$. This removes pressure modes with weak coupling to the retained velocity directions. The heuristic provides no proven lower bound on β_h; pressure certification requires a separate proven inf-sup constant (Section F).

F CERTIFICATE-STYLE MAJORANTS

Gram-matrix integration. For a Fourier feature $\phi_j(x) = \sin(\omega_j \cdot x + b_j)$ in d dimensions, we use

$$\phi_j(x) = \mathrm{Im}\left(e^{i(\omega_j \cdot x + b_j)}\right).$$

Products of features are sums of complex exponentials. Their box integrals factor into univariate closed forms:

$$\int_{[0,1]^d} e^{i\omega_j \cdot x} e^{-i\omega_k \cdot x}\, dx = \prod_{l=1}^{d} \int_0^1 e^{i(\omega_{j,l} - \omega_{k,l})x_l}\, dx_l = \prod_{l=1}^{d} \frac{e^{i(\omega_{j,l} - \omega_{k,l})} - 1}{i(\omega_{j,l} - \omega_{k,l})}.$$

We evaluate these expressions with stable `sinc`-type functions, using the limiting value when a frequency difference vanishes.

Derivative Gram matrices. Since $D^\alpha \phi_j = \omega_j^\alpha \sin(\omega_j \cdot x + b_j + |\alpha|\pi/2)$, we compute $\int (D^\alpha \phi_j)(D^\beta \phi_k)$ with the same closed forms, without quadrature or automatic differentiation.

Functional-dual assembly. We assemble each majorant as a quadratic form in the dual coefficients $\mathbf{s}$. For equation 11 with $\Sigma_h = \Sigma \mathbf{s}$, the Gram contributions are $\nu^{-1}(1+\varepsilon) \int \Sigma^\top \Sigma$ and $\nu^{-1}(1+\varepsilon^{-1}) C_\Omega^2 \int (\nabla \cdot \Sigma)^\top (\nabla \cdot \Sigma)$, together with the corresponding source terms. For $B_\varepsilon \succ 0$, the minimizer and majorant are

$$\mathbf{s}^\star = B_\varepsilon^{-1} \mathbf{b}_\varepsilon, \qquad \mathcal{M}_\varepsilon^\star = \mathrm{const} - \mathbf{b}_\varepsilon^\top B_\varepsilon^{-1} \mathbf{b}_\varepsilon,$$

followed by minimization over ε. Both residuals use the same dual field. Diffusion enters through $\nabla \cdot \Sigma_h$; the equilibrium residual contains no $\mathcal{L}\mathbf{u}_h$ term (Appendix B).

For the Maxwell row, the two-dimensional conventions are $\nabla \times E = \partial_x E_y - \partial_y E_x$ and $\nabla \times w = (\partial_y w, -\partial_x w)$, so $\nabla \times \nabla \times E = -\Delta E$ for divergence-free E. The displayed Poisson and Maxwell majorants bound the squared energy norms; their square roots give the energy-norm bounds used in effectivity ratios.

Table 10: Correspondence between each manufactured PDE, its fields, and the constant entering its majorant. The dual field is optional: Stokes uses a stress-like tensor Σ_h, Poisson a vector flux σ_h, and Maxwell a free scalar w; the energy norm is the one used to measure the reported errors and effectivity indices.

Problem, primal, and dual field	Energy norm, constant, squared majorant
Stokes, $\Omega = [0,1]^2$: primal $\mathbf{u}$ (zero trace), p (zero mean); dual Σ_h tensor in $H(\mathrm{Div};\Omega)$ with no condition on $\Sigma_h n$; $-\nu\Delta\mathbf{u}+\nabla p = \mathbf{f}$, $\nabla\cdot\mathbf{u} = 0$.	Energy $\nu^{1/2}\|\nabla e\|_{L^2}$, $e = \mathbf{u} - \mathbf{u}_h$; constant $C_\Omega = (\pi^2 \sum_l w_l^{-2})^{-1/2}$; majorant equation 11.
Poisson: primal u (zero trace); dual vector flux σ_h in $H(\mathrm{Div};\Omega)$; $-\Delta u = f$.	Energy $\|\nabla e\|_{L^2}$, $e = u - u_h$; same C_Ω; bound of Appendix F: $(1+\beta)\|\sigma_h - \nabla u_h\|_{L^2}^2 + (1+\beta^{-1})C_\Omega^2\|\nabla\cdot\sigma_h + f\|_{L^2}^2$.
Maxwell (TE): primal E (boundary conforming in the manufactured setting); dual free scalar $w \in H^1(\Omega)$, no boundary condition; $\alpha E + \nabla\times\nabla\times E = \mathbf{f}$, $\alpha = \kappa^2$.	Energy $\big(\alpha\|E - E_h\|_{L^2}^2 + \|\nabla\times(E-E_h)\|_{L^2}^2\big)^{1/2}$; constant α; majorant $\alpha^{-1}\|\mathbf{f} - \alpha E_h - \nabla\times w\|_{L^2}^2 + \|w - \nabla\times E_h\|_{L^2}^2$, minimized over w in the frozen sine basis.

Functional majorant (Poisson, scalar). For the Poisson experiments we use the same split form on the scalar field u, with the vector flux σ_h in $H(\nabla\cdot;\Omega)$ built from frozen sine gradients. For $e = u - u_h$ with zero trace, $\beta > 0$, and the same box Poincaré constant C_Ω,

$$\|\nabla e\|_{L^2}^2 \le (1+\beta)\|\sigma_h - \nabla u_h\|_{L^2}^2 + (1+\beta^{-1})C_\Omega^2\|\nabla\cdot\sigma_h + f\|_{L^2}^2 ,$$

and the right-hand side is minimized over the dual coefficients and β in closed form. The Dirichlet trace of e vanishes, so σ_h has no normal boundary condition. The Stokes functional majorant equation 11 uses the corresponding tensor-stress form.

Strong-residual majorant (Stokes). For Stokes we report the strong-residual/Poincaré bound

$$\|\mathbf{u} - \mathbf{u}_h\|_{\mathcal{A}} \le C_\Omega \nu^{-1/2}\|\mathbf{f} - \mathcal{L}\mathbf{u}_h\|_{L^2} .$$

As in (D0) of Proposition 3, this requires $\mathbf{u}_h \in H_0^1(\Omega)^d$ and continuous divergence-freedom. It uses the momentum residual without a dual field. The functional majorant in equation 11 instead uses a stress mismatch and an equilibrium residual, with $\Sigma_h \in H(\mathrm{Div};\Omega)$. Dirichlet velocity makes the error trace vanish, so the dual normal trace is unrestricted. Collocation constraints alone do not establish the admissibility required by either continuum bound.

Validity and arithmetic. The continuum Stokes bounds require exact primal boundary traces, continuous divergence-freedom, and a proven Poincaré constant; pressure bounds also require a proven continuous inf-sup constant. The functional form additionally requires dual $H(\mathrm{Div};\Omega)$ admissibility. The inequalities hold for the analytic quantities in exact arithmetic. Our float64 majorants and effectivity indices use no interval or outward-rounded arithmetic and do not certify roundoff. Poisson majorant sharpness depends on both the dual space and the split parameter; the effectivity reaches 4.158×10^3 in the fixed-parameter sweep.

Validation protocol. Table 19 tests the float64 inequality $M \ge \mathcal{E}$ between the assembled majorant M and the manufactured solution's energy error $\mathcal{E}$. We vary feature counts, interior and boundary sample counts, thresholds, and seeds. All 29/29 Poisson and 17/17 unanchored Maxwell settings satisfy $M \ge \mathcal{E}$. The last row of Table 19 adds 17 Maxwell settings whose dual basis contains the four frequencies of the exact manufactured flux. The optimal reconstruction recovers that flux, placing the majorant near $\mathcal{E}$ at float64 precision. The recorded comparison satisfies $M \ge \mathcal{E}$ in 16 of the 17 settings. Repeated runs give a largest relative deficit of 2.4×10^{-10}; allowing a relative tolerance of 5×10^{-10} gives 17/17. These near-equality comparisons are sensitive to floating-point tolerance and do not provide a roundoff enclosure.

G ADDITIONAL EXPERIMENTS AND ABLATIONS

Table 11: Advection coefficients and long-time stability (frozen integer Fourier basis, 7 frequency pairs). Panels (A)–(B) are SP-FNO; panel (C) lists HCS/SCS time-segmentation values from Roy & Castonguay (2024), as compiled in Datar et al. (2026).

(A) β sweep (final time $= 1$)

β	Rel. L^2 err. $\downarrow$	Strong res. RMS $\downarrow$
0.01	3.768×10^{-15}	3.433×10^{-18}
0.1	1.675×10^{-15}	3.462×10^{-17}
1	8.928×10^{-16}	4.003×10^{-16}
10	7.113×10^{-15}	3.890×10^{-15}
40	2.832×10^{-14}	1.561×10^{-14}
100	6.760×10^{-14}	3.876×10^{-14}
1000	6.068×10^{-13}	3.995×10^{-13}
10000	7.072×10^{-12}	3.894×10^{-12}

(B) Long-time (T=1000, β=1)

Final time	Rel. L^2 err. $\downarrow$	Max energy drift $\downarrow$
1000	6.415×10^{-13}	4.276×10^{-12}

(C) HCS/SCS time segmentation (T=1, β=40)

Method	c=30	c=100
HCS-PINN (n_t=4)	3.4179×10^{-3}	9.6380×10^{-1}
HCS-PINN (n_t=10)	2.1304×10^{-3}	6.2772×10^{-3}

Table 12: Stokes feature-count scaling (seed 2026).

# Features	$L^2(u) \downarrow$	$L^2(p) \downarrow$	PDE res. RMS $\downarrow$
32	0.9971	0.09092	0.7226
64	0.1077	0.04431	0.1602
128	2.548×10^{-5}	1.895×10^{-5}	2.429×10^{-4}
256	4.995×10^{-8}	6.225×10^{-8}	8.407×10^{-7}
512	6.909×10^{-8}	1.685×10^{-7}	1.565×10^{-6}

Table 13: Stokes sampling, ridge, boundary weight (seed 2026).

Factor	Setting	$L^2(u) \downarrow$	$L^2(p) \downarrow$	PDE res. RMS $\downarrow$
Interior pts	1000	3.532×10^{-5}	2.859×10^{-5}	2.864×10^{-4}
	5000	2.548×10^{-5}	1.895×10^{-5}	2.429×10^{-4}
	10000	2.417×10^{-5}	2.025×10^{-5}	2.389×10^{-4}
Ridge	0	2.852×10^{-5}	2.440×10^{-5}	2.638×10^{-4}
	$1e-12$	2.832×10^{-5}	2.416×10^{-5}	2.627×10^{-4}
	$1e-9$	2.548×10^{-5}	1.895×10^{-5}	2.429×10^{-4}
	$1e-6$	5.675×10^{-5}	4.457×10^{-5}	5.127×10^{-4}
Boundary weight	10	1.451×10^{-5}	1.476×10^{-5}	1.779×10^{-4}
	50	2.548×10^{-5}	1.895×10^{-5}	2.429×10^{-4}
	200	9.526×10^{-5}	7.130×10^{-5}	6.388×10^{-4}

Table 14: Stokes frequency distribution, solver backend, precision.

(A) Frequency distribution (F=128, seeds 2026–2028)

Metric	Gaussian	Student-t
$L^2(u)\downarrow$	$1.144 \times 10^{-5} \pm 8.913 \times 10^{-6}$	$6.187 \times 10^{-4} \pm 8.071 \times 10^{-4}$
$L^2(p)\downarrow$	$1.003 \times 10^{-5} \pm 6.313 \times 10^{-6}$	$5.604 \times 10^{-4} \pm 7.256 \times 10^{-4}$
PDE res. RMS $\downarrow$	$1.187 \times 10^{-4} \pm 8.788 \times 10^{-5}$	$5.220 \times 10^{-3} \pm 6.821 \times 10^{-3}$

(B) LSQ backend (seed 2026)

Backend	$L^2(u)\downarrow$	Div. RMS $\downarrow$	Solve time (s) $\downarrow$
CUDA gels	2.402×10^{-5}	7.586×10^{-15}	0.01941
CUDA normal eqns.	2.403×10^{-5}	7.597×10^{-15}	0.009682

(C) Numerical precision (seed 2026)

Precision	$L^2(u)\downarrow$	Div. RMS $\downarrow$	Solve time (s) $\downarrow$
float32	2.469×10^{-5}	3.831×10^{-6}	0.02987
float64	2.402×10^{-5}	7.586×10^{-15}	0.01949

Table 15: Manufactured-frequency prior ablation (paired seeds 0/1/2).

Metric	Prior off	Prior on
$L^2(u)\downarrow$	$4.091 \times 10^{-8} \pm 6.895 \times 10^{-9}$	$1.267 \times 10^{-8} \pm 1.179 \times 10^{-9}$
$L^2(v)\downarrow$	$4.695 \times 10^{-8} \pm 7.605 \times 10^{-9}$	$1.343 \times 10^{-8} \pm 1.370 \times 10^{-9}$
$L^2(p)\downarrow$	$6.330 \times 10^{-8} \pm 1.697 \times 10^{-9}$	$4.363 \times 10^{-8} \pm 8.098 \times 10^{-9}$
PDE res. RMS $\downarrow$	$8.494 \times 10^{-7} \pm 7.331 \times 10^{-8}$	$3.761 \times 10^{-7} \pm 4.952 \times 10^{-8}$

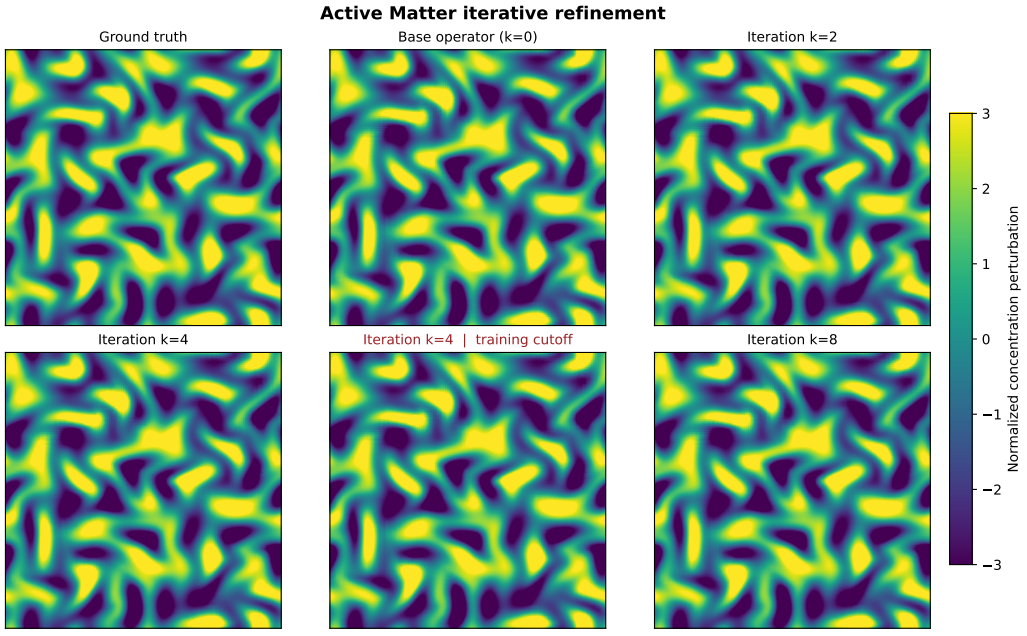

Figure 6: Active-matter iterative refinement (shared weight, The Well test set). Left to right, top to bottom: ground truth, base operator (k=0), k=2 and k=4 refinements, the marked K=4 training cutoff, and k=8 extrapolation on a shared normalized-concentration colormap. With K=4 steps, SP-FNO+refinement reaches 0.02733 VRMSE; Table 8 gives the literature comparison and Appendix D describes the evaluation budgets.

Table 16: Navier–Stokes robustness sweeps (seed 2026).

(A) Feature count

# Features	$L^2(u)\downarrow$	$L^2(v)\downarrow$	$L^2(p)\downarrow$	PDE res. RMS $\downarrow$
32	0.01080	0.01195	0.1396	0.1093
64	1.232×10^{-4}	1.389×10^{-4}	2.063×10^{-3}	1.560×10^{-3}
96	1.693×10^{-3}	1.739×10^{-3}	0.03496	0.01319

(B) Viscosity

ν	$L^2(u)\downarrow$	$L^2(v)\downarrow$	$L^2(p)\downarrow$	PDE res. RMS $\downarrow$
0.1	1.390×10^{-4}	1.216×10^{-4}	4.566×10^{-3}	0.03274
0.05	3.249×10^{-4}	2.897×10^{-4}	5.549×10^{-3}	0.02800
0.01	1.693×10^{-3}	1.739×10^{-3}	0.03496	0.01319

(C) Newton vs. Picard

Solver	$L^2(u)\downarrow$	$L^2(v)\downarrow$	$L^2(p)\downarrow$	PDE res. RMS $\downarrow$
Newton	1.693×10^{-3}	1.739×10^{-3}	0.03496	0.01319
Picard	0.6179	0.6192	7.211	3.022

Table 17: Active-matter refinement trajectory.

Variant	K	VRMSE $\downarrow$	RMSE $\downarrow$	Div. RMS $\downarrow$
Proj. train + proj. eval	0	3.082	0.1774	4.520×10^{-7}
	1	2.644 ± 0.2144	0.09907 ± 0.001792	$1.871 \times 10^{-6} \pm 6.363 \times 10^{-8}$
	2	2.636 ± 0.5324	0.08105 ± 0.002681	$2.741 \times 10^{-6} \pm 1.956 \times 10^{-7}$
	3	2.715 ± 0.7669	0.08009 ± 0.004296	$2.831 \times 10^{-6} \pm 9.547 \times 10^{-8}$
	4	2.849 ± 0.9447	0.08394 ± 0.005849	$2.927 \times 10^{-6} \pm 5.602 \times 10^{-8}$
	5	3.093 ± 1.022	0.08973 ± 0.006682	$2.987 \times 10^{-6} \pm 4.043 \times 10^{-8}$
Proj. train, no proj. eval	0	3.131	0.1778	0.2804
	1	4.059 ± 0.6389	0.1127 ± 0.004865	1.223 ± 0.2911
	2	3.973 ± 0.7936	0.09936 ± 0.009407	1.668 ± 0.8244
	3	4.534 ± 1.355	0.1035 ± 0.01608	1.816 ± 0.9357
	4	4.733 ± 1.728	0.1098 ± 0.02091	1.909 ± 0.9522
	5	5.133 ± 1.944	0.1181 ± 0.02409	2.026 ± 0.9437
No proj. train + eval	0	3.131	0.1778	0.2804
	1	2.692 ± 0.2102	0.1018 ± 0.001502	0.8926 ± 0.09514
	2	2.794 ± 0.4955	0.08124 ± 0.002031	1.059 ± 0.02392
	3	2.752 ± 0.5961	0.07797 ± 0.002262	0.9869 ± 0.06262
	4	2.778 ± 0.6589	0.07983 ± 0.003229	0.9323 ± 0.08835
	5	2.908 ± 0.6805	0.08483 ± 0.003798	0.9055 ± 0.1066

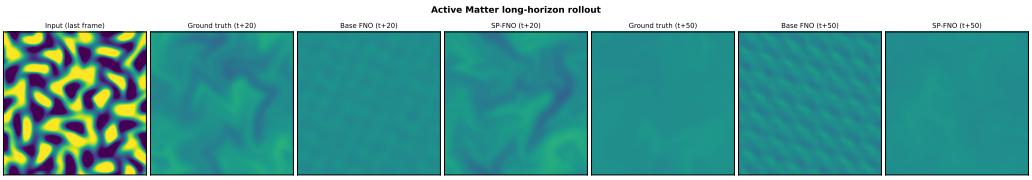

Figure 7: Active-matter long-horizon autoregressive rollout. Left: last input frame; then ground truth, base FNO, and SP-FNO at $t+20$ and $t+50$. The concentration panels show changes in contrast and spatial patterns; Table 4D reports aggregate rollout errors.

Table 18: Inf-sup filtering threshold scan and coupling-strength scan.

(A) Filter threshold (mixed Stokes)

Threshold	#Pressure dims	$\beta_h \uparrow$	Schur cond. $\downarrow$	$L^2(p) \downarrow$
0	7	1.308×10^{-4}	8.553×10^{9}	2.698×10^{4}
10^{-8}	7	1.308×10^{-4}	8.553×10^{9}	2.698×10^{4}
10^{-6}	7	1.308×10^{-4}	8.553×10^{9}	2.698×10^{4}
10^{-5}	7	1.308×10^{-4}	8.553×10^{9}	2.698×10^{4}
10^{-4}	6	**1.688**	**51.34**	**1.337**
10^{-3}	6	1.688	51.34	1.337
10^{-2}	6	1.688	51.34	1.337

(B) Coupling strength (whitened inf-sup)

γ	Raw monolithic cond. $\downarrow$	Block-scaled cond. $\downarrow$	Projected SPD cond. $\downarrow$
1	8.130	8.130	1.805
10	14.40	5.211	1.805
100	136.2	14.40	1.805
1000	1355	43.65	1.805
10000	1.354×10^{4}	136.2	1.805

Table 19: Functional-majorant sweep summary.

Problem	Sweep	# settings	Effectivity index	Checks $M \geq \mathcal{E}$
Mfg. Poisson	primal convergence	7	1.000–1.000	7/7
Mfg. Poisson	dual augmentation	8	1.000–9.252	8/8
Mfg. Poisson	fixed β	13	4.914–4.158×10^{3}	13/13
Mfg. Poisson	analytic alternating β	1	4.889	1/1
Maxwell (no anchor)	primal/seed	12	1.783–11.98	12/12
Maxwell (no anchor)	dual dim	5	7.457–14.51	5/5
Maxwell (mfg. flux prior)	(diag.)	17	1.000–1.000	16/17

Table 20: $\beta = 40$ RMSE multi-metric reference.

Method	Year	RMSE $\downarrow$	Rel. $L^2 \downarrow$	Time (s) $\downarrow$
PINN (L-BFGS, lr=0.1, d=10)	2019	3.09×10^{-3}	4.37×10^{-3}	207.25
PINN (L-BFGS, lr=0.01, d=10)	2019	1.46×10^{-2}	2.07×10^{-2}	205.08
PINN (Adam, lr=0.001, d=10)	2019	3.66×10^{-1}	5.18×10^{-1}	168.14
SP-FNO (ours)	2026	$\mathbf{2.0029 \times 10^{-14}}$	$\mathbf{2.8324 \times 10^{-14}}$	$\mathbf{1.1870 \times 10^{-2}}$

Table 21: Scalar Helmholtz 2D and Poisson 5D.

(A) CPU

Metric	Helmholtz 2D	Poisson 5D
Rel. L^2 err. $\downarrow$	$2.585 \times 10^{-2} \pm 1.519 \times 10^{-2}$	$6.246 \times 10^{-1} \pm 4.124 \times 10^{-2}$
RMSE $\downarrow$	$1.251 \times 10^{-2} \pm 7.419 \times 10^{-3}$	$2.044 \pm 1.368 \times 10^{-1}$
PDE res. RMS $\downarrow$	$5.339 \times 10^{-1} \pm 3.296 \times 10^{-1}$	$6.055 \times 10^{1} \pm 3.395$
Time (s) $\downarrow$	$3.335 \times 10^{-1} \pm 1.191 \times 10^{-1}$	$5.673 \times 10^{-1} \pm 4.052 \times 10^{-2}$

(B) RTX 4090 (float64, F=1500, 10^4 interior)

Metric	Helmholtz 2D	Poisson 5D
Rel. L^2 err. $\downarrow$	$3.514 \times 10^{-9} \pm 6.249 \times 10^{-10}$	$1.028 \times 10^{-7} \pm 1.751 \times 10^{-8}$
RMSE $\downarrow$	$1.692 \times 10^{-9} \pm 2.923 \times 10^{-10}$	$3.353 \times 10^{-7} \pm 5.684 \times 10^{-8}$
PDE res. RMS $\downarrow$	$2.089 \times 10^{-7} \pm 6.171 \times 10^{-8}$	$1.571 \times 10^{-5} \pm 1.747 \times 10^{-6}$
Total time (s) $\downarrow$	0.2043 ± 0.0916	0.2058 ± 0.0813

Table 22: 5D nonlinear reaction–diffusion. Panel A quotes the published baselines from Datar et al. (2026); panel B reports SP-FNO.

(A) Published baselines

Method	Time (s)	RMSE ↓	Rel. L^2 ↓
PINN (ADAM)	171.43	$1.25 \times 10^{-1} \pm 6.60 \times 10^{-3}$	$3.40 \times 10^{-1} \pm 1.79 \times 10^{-2}$
PINN (LBFGS)	183.38	$3.33 \times 10^{-2} \pm 1.54 \times 10^{-2}$	$3.33 \times 10^{-2} \pm 1.54 \times 10^{-2}$
Frozen-PINN-elm	621.2	$6.17 \times 10^{-3} \pm 2.02 \times 10^{-4}$	$1.67 \times 10^{-2} \pm 5.49 \times 10^{-4}$
Frozen-PINN-swim	117.24	$2.09 \times 10^{-3} \pm 1.91 \times 10^{-5}$	$5.70 \times 10^{-3} \pm 5.19 \times 10^{-5}$
Frozen-PINN-swim (proj.)	12.43	$3.67 \times 10^{-5} \pm 2.28 \times 10^{-9}$	$9.99 \times 10^{-5} \pm 6.21 \times 10^{-9}$

(B) SP-FNO (RTX 4090, 3 seeds)

Metric	**SP-FNO (ours)**
Time (s) ↓	0.2658 ± 0.1460
RMSE ↓	$\mathbf{1.116 \times 10^{-5} \pm 6.896 \times 10^{-7}}$
Rel. L^2 ↓	$\mathbf{1.695 \times 10^{-5} \pm 1.047 \times 10^{-6}}$

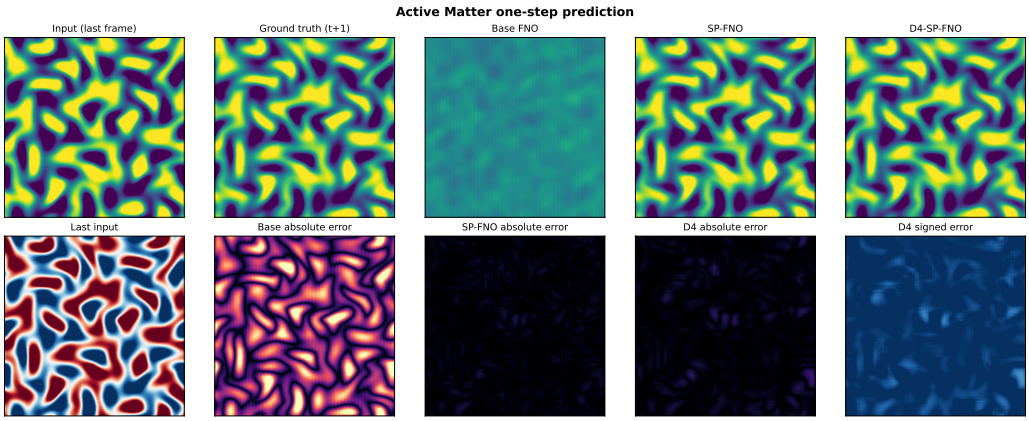

Figure 8: One-step active-matter prediction ($4 \rightarrow 1$, The Well test set). Top: last input frame, ground truth, base FNO, SP-FNO, and D4-SP-FNO on a shared normalized-concentration colormap. Bottom: input, absolute errors of FNO / SP-FNO / D4-SP-FNO, and the signed error of D4-SP-FNO. The learned FNO baseline loses the filamentary structure, while SP-FNO preserves the sharp orientation, and the D4 ensemble reduces error in this example.

Table 23: Non-Fourier/OOD initial conditions (3 seeds; 16 frequency pairs, a $\{1, \ldots, 16\}$ integer Fourier ladder). Initial conditions: multimodal $u_0(x) = \sin x + 0.5\sin(3x)$ (in basis), high frequency $u_0(x) = \sin(9x) + 0.5\sin(17x)$ (one harmonic outside the ladder), wave packet $u_0(x) = \exp\big(-\frac{1}{2}[\sin(\frac{x-\pi}{2})/0.24]^2\big)\cos(7(x-\pi))$ (broad spectrum, partly outside the ladder).

(A) multimodal	β=40	β=100
Rel. L^2 ↓	$6.311 \times 10^{-14} \pm 4.740 \times 10^{-17}$	$2.005 \times 10^{-13} \pm 1.213 \times 10^{-16}$
RMSE ↓	$4.990 \times 10^{-14} \pm 3.748 \times 10^{-17}$	$1.585 \times 10^{-13} \pm 9.589 \times 10^{-17}$
Strong res. RMS ↓	$8.204 \times 10^{-14} \pm 9.201 \times 10^{-15}$	$2.045 \times 10^{-13} \pm 2.273 \times 10^{-14}$
Period. mismatch ↓	$1.217 \times 10^{-15} \pm 6.542 \times 10^{-18}$	$1.221 \times 10^{-15} \pm 0$
(B) high frequency	β=40	β=100
Rel. L^2 ↓	$4.472 \times 10^{-1} \pm 0$	$4.472 \times 10^{-1} \pm 5.551 \times 10^{-17}$
RMSE ↓	$3.536 \times 10^{-1} \pm 5.551 \times 10^{-17}$	$3.536 \times 10^{-1} \pm 5.551 \times 10^{-17}$
Strong res. RMS ↓	$8.139 \times 10^{-13} \pm 1.857 \times 10^{-13}$	$2.036 \times 10^{-12} \pm 4.819 \times 10^{-13}$
Period. mismatch ↓	$2.146 \times 10^{-15} \pm 1.019 \times 10^{-15}$	$2.146 \times 10^{-15} \pm 1.108 \times 10^{-15}$
(C) wave packet	β=40	β=100
Rel. L^2 ↓	$3.957 \times 10^{-3} \pm 3.087 \times 10^{-18}$	$3.957 \times 10^{-3} \pm 2.929 \times 10^{-17}$
RMSE ↓	$1.037 \times 10^{-3} \pm 7.192 \times 10^{-19}$	$1.037 \times 10^{-3} \pm 7.628 \times 10^{-18}$
Strong res. RMS ↓	$2.006 \times 10^{-13} \pm 1.133 \times 10^{-14}$	$4.971 \times 10^{-13} \pm 2.809 \times 10^{-14}$
Period. mismatch ↓	$1.961 \times 10^{-15} \pm 3.735 \times 10^{-16}$	$2.040 \times 10^{-15} \pm 2.891 \times 10^{-16}$

Table 24: 2D Maxwell/TE multi-seed stability.

Metric	**SP-FNO (ours)**
Rel. L^2 err. ↓	1.500×10^{-9} $\pm 5.579 \times 10^{-10}$ 6.939×10^{-8}
PDE res. RMS ↓	$\pm 2.530 \times 10^{-8}$ 3.056×10^{-16}
Div. max ↓	$\pm 3.970 \times 10^{-17}$ 0.2270
Time (s) ↓	$\pm$0.1283

Table 25: Same-protocol backbone comparison.

Metric	Frozen FNO (2026)	**SP-FNO (ours)**
VRMSE ↓	3.1307	**2.715** **$\pm$0.7669**
RMSE ↓	0.17777	**0.08009** **± 0.004296**
Div. RMS ↓	2.804×10^{-1}	**2.831×10^{-6}** **$\pm 9.547 \times 10^{-8}$**

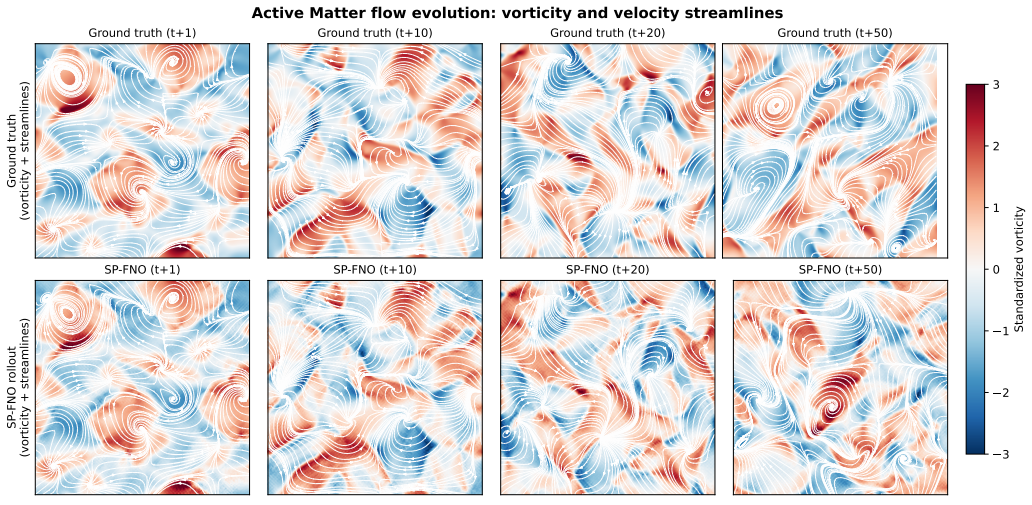

Figure 9: Active-matter flow evolution (sample 25). Ground truth (top) and SP-FNO rollout (bottom) at $t{+}1, t{+}10, t{+}20, t{+}50$. Vortices, shear bands, and filaments persist, with increasing phase differences at longer horizons.