Supplementary Information

S1 Error analysis of Geometry-aware LegONet

This section gives a standard stability statement for the boundary-adapted realization used in Geometry-aware LegONet. The purpose is to make explicit how the rollout error separates into three standard contributions: approximation by the boundary-adapted space, mismatch of the frozen realized mechanism responses, and time-integration error.

Let $u(t)$ be the embedded-domain solution on Ω over $t \in [0, T]$, after the affine map from Ω_{phys} to $\Omega \subset Q$. Throughout this section, $\mathbf{x} = (x, y)$ denotes the embedded spatial coordinate.

Definition S1.1 (Boundary-adapted approximation error)**.** Let $\Omega \subset Q$ be an embedded target domain equipped with the boundary-adapted coordinate representation $\mathbf{a} = \mathbf{a}_{\mathrm{bc}}(t) + N_\Omega \mathbf{z}$. For a target trajectory $u(t)$ on $t \in [0, T]$, define

$$\eta_r(T) = \sup_{0 \le t \le T} \inf_{\mathbf{z} \in \mathbb{R}^r} \left\| u(t) - \mathcal{R}_Q\big(\mathbf{a}_{\mathrm{bc}}(t) + N_\Omega \mathbf{z}\big)\big|_\Omega \right\|_{L^2(\Omega)}.$$

We call $\eta_r(T)$ the boundary-adapted best-approximation error.

The quantity $\eta_r(T)$ measures the expressive power of the reduced boundary-adapted space selected by N_Ω. It is the smallest $L^2(\Omega)$ error that can be achieved by any coefficient vector satisfying the sampled boundary constraints. The size of $\eta_r(T)$ is controlled by three factors: the approximation power of the ambient trial space $\mathcal{V}_K(Q)$, the regularity of an ambient extension of the target solution, and the stability of the sampled trace constraints used to construct N_Ω. Increasing the Fourier cutoff K enriches the ambient trial space, while refining the boundary samples improves the sampled representation of the boundary trace. However, these refinements are useful only when the sampled boundary matrix C and the Gram matrix $Z^\top M_\Omega Z$ remain sufficiently well conditioned for a stable numerical null-space construction. Therefore, the approximation error is not claimed to decrease monotonically with K or with the number of boundary samples; it is a problem-dependent term that reflects ambient resolution, boundary sampling, solution regularity and numerical conditioning.

In the experiments, this is why boundary accuracy is checked on independent boundary points not used to construct C. The reported boundary residuals therefore test whether the sampled boundary-adapted space generalizes from the construction points to the full target boundary.

Assumption S1.1 (Stability and approximation setting)**.** *Let $u(t)$ be the embedded-domain solution on Ω over $t \in [0,T]$, and assume that its trajectory remains in a bounded subset of the solution space. Suppose the target PDE is decomposed into mechanisms as*

$$\partial_t u = \sum_{i=1}^{N_{\mathrm{blk}}} c_i L_i^{\Omega}(u),$$

and let $\mathbf{q}_i^{\Omega,\star}$ denote the exact target-domain Galerkin response of L_i^{Ω} in the boundary-adapted coordinates. Let $\mathbf{q}_i^{\Omega}$ denote the corresponding realized response obtained from the frozen ambient library and the deterministic target-domain assembly. Let $\mathbf{z}_{\mathrm{gal}}(t)$ be the exact Galerkin trajectory driven by $\{\mathbf{q}_i^{\Omega,\star}\}_{i=1}^{N_{\mathrm{blk}}}$, and let $\mathbf{z}_\theta(t)$ be the continuous-time trajectory driven by $\{\mathbf{q}_i^{\Omega}\}_{i=1}^{N_{\mathrm{blk}}}$. We assume the following standard stability and approximation conditions.

1. *The exact reduced vector field is Lipschitz on the relevant bounded set, with Lipschitz constant $L_r > 0$.*
2. *With $\eta_r(T)$ defined in Definition S1.1, the exact Galerkin trajectory satisfies the quasi-optimality bound*

$$\sup_{0\le t\le T} \left\| u(t) - \mathcal{R}_Q\left(\mathbf{a}_{\mathrm{bc}}(t) + N_\Omega \mathbf{z}_{\mathrm{gal}}(t)\right)\big|_\Omega \right\|_{L^2(\Omega)} \le C_{\mathrm{gal}} \eta_r(T),$$

   where $C_{\mathrm{gal}} > 0$ is independent of the learned-response error.
3. *The frozen realized responses satisfy the uniform reduced-response perturbation bound*

$$\sup_{\mathbf{z},t} \left\| \sum_{i=1}^{N_{\mathrm{blk}}} c_i \left[\mathbf{q}_i^{\Omega}(\mathbf{z},t) - \mathbf{q}_i^{\Omega,\star}(\mathbf{z},t) \right] \right\|_2 \le \varepsilon_\theta,$$

   with $\varepsilon_\theta \ge 0$.
4. *The chosen time integrator has order p and its global discretization error on the learned reduced system is bounded by $C_{\mathrm{time}}\Delta t^p$, with $C_{\mathrm{time}} > 0$.*

These assumptions are not specific to a particular neural architecture. They are the standard ingredients for a Galerkin-reduced dynamical system with a perturbed vector field: stability of the reduced flow, quasi-optimality of the boundary-adapted trial space, uniform reduced-response approximation error, and the usual global time-integration error.

Theorem S1.1 (Error decomposition for Geometry-aware LegONet rollout)**.** *Let $0 = t_0 < t_1 < \cdots < t_{N_T} = T$ be the rollout time grid, with $\Delta t = \max_{0\le n<N_T}(t_{n+1} - t_n)$. Under Assumption S1.1, let $\mathbf{z}_\theta^n$ be the fully discrete reduced state produced by Geometry-aware LegONet at time t_n, and define*

$$\mathbf{a}_\theta^n = \mathbf{a}_{\mathrm{bc}}(t_n) + N_\Omega \mathbf{z}_\theta^n, \qquad u_\theta^n = \mathcal{R}_Q \mathbf{a}_\theta^n|_\Omega .$$

Then

$$\max_{0\le n\le N_T} \|u(t_n) - u_\theta^n\|_{L^2(\Omega)} \le C_{\mathrm{gal}}\eta_r(T) + C_\Omega e^{L_r T}\left(\|\mathbf{z}_\theta^0 - \mathbf{z}_{\mathrm{gal}}(0)\|_2 + T\varepsilon_\theta\right) + C_{\mathrm{time}}\Delta t^p.$$

Here $C_\Omega > 0$ accounts for passing from the discrete M_Ω-normalized reduced-coordinate norm to the continuous $L^2(\Omega)$ norm. Under an affine map from Ω_{phys} to Ω, the same estimate holds on Ω_{phys} up to fixed geometry-dependent constants.

Proof Let

$$\mathbf{a}_{\mathrm{gal}}(t) = \mathbf{a}_{\mathrm{bc}}(t) + N_\Omega \mathbf{z}_{\mathrm{gal}}(t), \qquad u_{\mathrm{gal}}(t) = \mathcal{R}_Q \mathbf{a}_{\mathrm{gal}}(t)\big|_\Omega,$$

where $\mathbf{z}_{\mathrm{gal}}(t)$ is the exact reduced Galerkin trajectory. Let $\widetilde{\mathbf{z}}_\theta(t)$ be the continuous-time trajectory of the learned reduced system with $\widetilde{\mathbf{z}}_\theta(0) = \mathbf{z}_\theta^0$, and set

$$\widetilde{\mathbf{a}}_\theta(t) = \mathbf{a}_{\mathrm{bc}}(t) + N_\Omega \widetilde{\mathbf{z}}_\theta(t), \qquad \widetilde{u}_\theta(t) = \mathcal{R}_Q \widetilde{\mathbf{a}}_\theta(t)\big|_\Omega.$$

The fully discrete Geometry-aware LegONet rollout is denoted by $\mathbf{z}_\theta^n$, with

$$\mathbf{a}_\theta^n = \mathbf{a}_{\mathrm{bc}}(t_n) + N_\Omega \mathbf{z}_\theta^n, \qquad u_\theta^n = \mathcal{R}_Q \mathbf{a}_\theta^n\big|_\Omega.$$

For each t_n, the triangle inequality gives

$$\begin{aligned}\|u(t_n) - u_\theta^n\|_{L^2(\Omega)} \le \|u(t_n) - u_{\mathrm{gal}}(t_n)\|_{L^2(\Omega)} + \|u_{\mathrm{gal}}(t_n) - \widetilde{u}_\theta(t_n)\|_{L^2(\Omega)} \\ + \|\widetilde{u}_\theta(t_n) - u_\theta^n\|_{L^2(\Omega)}.\end{aligned}$$

By Assumption S1.1 and Definition S1.1,

$$\sup_{0\le t\le T} \|u(t) - u_{\mathrm{gal}}(t)\|_{L^2(\Omega)} \le C_{\mathrm{gal}}\eta_r(T), \qquad C_{\mathrm{gal}} > 0.$$

We next bound the perturbation introduced by the frozen realized mechanism responses. Define the exact and learned reduced vector fields $\mathbf{f}_{\mathrm{gal}}, \mathbf{f}_\theta : \mathbb{R}^r \times [0,T] \to \mathbb{R}^r$ by

$$\mathbf{f}_{\mathrm{gal}}(\mathbf{z},t) = \sum_{i=1}^{N_{\mathrm{blk}}} c_i \mathbf{q}_i^{\Omega,\star}(\mathbf{z},t) - N_\Omega^\top M_\Omega \dot{\mathbf{a}}_{\mathrm{bc}}(t),$$

$$\mathbf{f}_\theta(\mathbf{z},t) = \sum_{i=1}^{N_{\mathrm{blk}}} c_i \mathbf{q}_i^{\Omega}(\mathbf{z},t) - N_\Omega^\top M_\Omega \dot{\mathbf{a}}_{\mathrm{bc}}(t).$$

Thus

$$\dot{\mathbf{z}}_{\mathrm{gal}}(t) = \mathbf{f}_{\mathrm{gal}}(\mathbf{z}_{\mathrm{gal}}(t),t), \qquad \dot{\widetilde{\mathbf{z}}}_\theta(t) = \mathbf{f}_\theta(\widetilde{\mathbf{z}}_\theta(t),t).$$

Let

$$\mathbf{e}(t) = \widetilde{\mathbf{z}}_\theta(t) - \mathbf{z}_{\mathrm{gal}}(t).$$

Then

$$\begin{aligned}\dot{\mathbf{e}}(t) &= \mathbf{f}_\theta(\widetilde{\mathbf{z}}_\theta(t),t) - \mathbf{f}_{\mathrm{gal}}(\mathbf{z}_{\mathrm{gal}}(t),t) \\ &= \left[\mathbf{f}_{\mathrm{gal}}(\widetilde{\mathbf{z}}_\theta(t),t) - \mathbf{f}_{\mathrm{gal}}(\mathbf{z}_{\mathrm{gal}}(t),t)\right] + \left[\mathbf{f}_\theta(\widetilde{\mathbf{z}}_\theta(t),t) - \mathbf{f}_{\mathrm{gal}}(\widetilde{\mathbf{z}}_\theta(t),t)\right].\end{aligned}$$

Using the Lipschitz bound for $\mathbf{f}_{\mathrm{gal}}$, with $L_r > 0$, and the uniform reduced-response error, with $\varepsilon_\theta \ge 0$, we obtain

$$\frac{d}{dt}\|\mathbf{e}(t)\|_2 \le L_r\|\mathbf{e}(t)\|_2 + \varepsilon_\theta.$$

By Gronwall's inequality [1,2], for $0 \le t \le T$,

$$\begin{aligned}\|\mathbf{e}(t)\|_2 &\le e^{L_r t}\|\mathbf{e}(0)\|_2 + \int_0^t e^{L_r(t-s)}\varepsilon_\theta\, ds \\ &\le e^{L_r t}\left(\|\mathbf{e}(0)\|_2 + t\varepsilon_\theta\right).\end{aligned}$$

Since $\mathbf{e}(0) = \mathbf{z}_\theta^0 - \mathbf{z}_{\mathrm{gal}}(0)$, this gives

$$\sup_{0\le t\le T} \left\|\widetilde{\mathbf{z}}_\theta(t) - \mathbf{z}_{\mathrm{gal}}(t)\right\|_2 \le e^{L_r T}\left(\|\mathbf{z}_\theta^0 - \mathbf{z}_{\mathrm{gal}}(0)\|_2 + T\varepsilon_\theta\right).$$

The reconstruction satisfies the continuity estimate

$$\left\|\mathcal{R}_Q N_\Omega(\mathbf{z}_1 - \mathbf{z}_2)\big|_\Omega\right\|_{L^2(\Omega)} \le C_\Omega \|\mathbf{z}_1 - \mathbf{z}_2\|_2, \qquad C_\Omega > 0,$$

for all $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^r$. The constant C_Ω accounts for the passage from the discrete M_Ω-norm to the continuous $L^2(\Omega)$-norm. Therefore,

$$\begin{aligned}\sup_{0\le t\le T} \|u_{\mathrm{gal}}(t) - \widetilde{u}_\theta(t)\|_{L^2(\Omega)} &\le C_\Omega \sup_{0\le t\le T} \left\|\mathbf{z}_{\mathrm{gal}}(t) - \widetilde{\mathbf{z}}_\theta(t)\right\|_2 \\ &\le C_\Omega e^{L_r T}\left(\|\mathbf{z}_\theta^0 - \mathbf{z}_{\mathrm{gal}}(0)\|_2 + T\varepsilon_\theta\right).\end{aligned}$$

By the assumed global error estimate for the chosen time integrator, there exist $C_{\mathrm{time}} > 0$ and $p \ge 1$ such that

$$\max_{0\le n\le N_T} \left\|\widetilde{u}_\theta(t_n) - u_\theta^n\right\|_{L^2(\Omega)} \le C_{\mathrm{time}} \Delta t^p.$$

Combining the three estimates yields

$$\begin{aligned}\max_{0\le n\le N_T} \|u(t_n) - u_\theta^n\|_{L^2(\Omega)} \le C_{\mathrm{gal}}\eta_r(T) + C_\Omega e^{L_r T}\left(\|\mathbf{z}_\theta^0 - \mathbf{z}_{\mathrm{gal}}(0)\|_2 + T\varepsilon_\theta\right) \\ + C_{\mathrm{time}}\Delta t^p.\end{aligned}$$

Finally, let $T_\Omega : \Omega_{\mathrm{phys}} \to \Omega$ be the affine embedding. Since its Jacobian is constant and nonsingular, there exist constants $0 < c_T \le C_T < \infty$, depending only on T_Ω, such that for every $v \in L^2(\Omega_{\mathrm{phys}})$,

$$c_T \|v\|_{L^2(\Omega_{\mathrm{phys}})} \le \|v \circ T_\Omega^{-1}\|_{L^2(\Omega)} \le C_T \|v\|_{L^2(\Omega_{\mathrm{phys}})}.$$

Hence the embedded-domain estimate transfers to Ω_{phys} up to fixed geometry-dependent constants. This completes the proof. □

Remark S1.1 (Interpretation of the error decomposition). The theorem separates the rollout error into geometry-transfer, mechanism-approximation and time-discretization components. The term $\eta_r(T)$ is set by the boundary-adapted trial space; the term ε_θ is set by the accuracy of the frozen realized mechanism responses; and the term Δt^p is set by the rollout integrator. The estimate therefore identifies the principal levers for improving accuracy: richer admissible coordinates, more accurate pretrained mechanism blocks and a finer or higher-order time integrator.

S2 Ambient block pretraining and transfer details

This section specifies the sampling distributions, block parameterizations, supervised objectives and target-domain realization used in the numerical experiments. All neural blocks are pretrained on the fixed ambient square $Q = [-1, 1]^2$, independently of the target domain, boundary condition and downstream equation. Throughout this section, $\mathbf{x} = (x, y) \in Q$ denotes the ambient spatial coordinate.

S2.1 Ambient Fourier space and coefficient prior

For a Fourier cutoff K, define $\mathcal{I}_K = \{(k,\ell) \in \mathbb{Z}^2 : \ k^2+\ell^2 \le K^2\}$, and let $\mathcal{I}_K^+ = \{(k,\ell) \in \mathcal{I}_K : \ k>0 \text{ or } (k=0,\ell>0)\}$ select one representative from each nonzero pair (k,ℓ) and $(-k,-\ell)$. Throughout this section, $\mathbf{x} = (x,y) \in Q$. The real Fourier space $\mathcal{V}_K(Q) = \operatorname{span}\{\phi_1,\dots,\phi_M\}$ contains the constant mode and

$$\sqrt{2}\cos\big(\pi(kx+\ell y)\big), \qquad \sqrt{2}\sin\big(\pi(kx+\ell y)\big), \qquad (k,\ell) \in \mathcal{I}_K^+.$$

The basis is orthonormal under $\langle u,v\rangle_Q = |Q|^{-1}\int_Q uv\,d\mathbf{x}$. For $K=22$, $M = 1+2|\mathcal{I}_K^+| = |\mathcal{I}_K| = 1517$.

A coefficient vector $\mathbf{a} \in \mathbb{R}^M$ represents

$$u_{\mathbf{a}}(\mathbf{x}) = \mathcal{R}_Q\mathbf{a} = \sum_{j=1}^{M} a_j\phi_j(\mathbf{x}) = \Phi_Q(\mathbf{x})\mathbf{a},$$

where $\Phi_Q(\mathbf{x}) = (\phi_1(\mathbf{x}),\dots,\phi_M(\mathbf{x}))$. On an ambient quadrature grid $\{\mathbf{x}_q^Q\}_{q=1}^{N_Q}$, the evaluation matrix $\Phi_Q^{\mathrm{quad}} \in \mathbb{R}^{N_Q\times M}$ is defined by $(\Phi_Q^{\mathrm{quad}})_{qj} = \phi_j(\mathbf{x}_q^Q)$.

Training coefficients are sampled independently from the diagonal Gaussian prior

$$a_j \sim \mathcal{N}(0,\sigma_j^2), \qquad \sigma_j = \Big(1+\sqrt{k_j^2+\ell_j^2}\Big)^{-1/2},$$

where (k_j,ℓ_j) is the wave vector associated with ϕ_j. This spectral weighting suppresses high-frequency coefficients relative to low-frequency coefficients and defines the finite-resolution amplitude regime used for block pretraining and held-out testing.

The generator form

$$\mathbf{F}_i^\theta(\mathbf{a}) = -G_i\nabla_{\mathbf{a}}E_i^\theta(\mathbf{a}) + J_i\nabla_{\mathbf{a}}H_i^\theta(\mathbf{a}) + \mathbf{R}_i(\mathbf{a})$$

is a structural template; each practical block remains single purpose. Dissipative mechanisms use an E-block, conservative or transport mechanisms use an H-block, and forcing, reaction or closure effects may be represented by an $\mathbf{R}$-block. This design follows the original LegONet framework[3].

All supervised data are generated on the ambient square Q, without target geometry or boundary information. The supervised quantity is chosen according to what must remain reusable after transfer. If a complete ambient coefficient response is transferable, the block is supervised directly against $\mathbf{F}_i(\mathbf{a}) = \mathcal{P}_Q[L_i^Q(\mathcal{R}_Q\mathbf{a})]$, evaluated analytically or by a trusted high-order reference discretization. If a mechanism must be reassembled weakly after the target domain is specified, as for nonlinear transport, we supervise the geometry-independent local constituent instead. Thus the training target is mechanism dependent, but the principle is the same: each block learns the most transferable representation of a physical mechanism on Q.

After pretraining, every library entry supplies an ambient mechanism that can be realized on a target domain through the boundary-adapted Galerkin interface. For a

target domain Ω, the online object used by the reduced dynamics is always the reduced response $\mathbf{q}_i^{\Omega}$. Geometry and boundary conditions change only the deterministic realization layer $(C, \mathbf{a}_{\mathrm{bc}}, N_\Omega, M_\Omega)$; the neural parameters of the ambient library remain fixed.

S2.2 Diffusion block

The exact ambient coefficient response of the Laplacian is

$$\mathbf{F}_\Delta(\mathbf{a}) = \mathcal{P}_Q\left[\Delta(\mathcal{R}_Q\mathbf{a})\right] = D_\Delta\mathbf{a},$$

where $D_\Delta \in \mathbb{R}^{M\times M}$ is diagonal. Following the dissipative LegONet template[3], the learned block is

$$\mathbf{F}_\Delta^\theta(\mathbf{a}) = -G_\Delta\nabla_{\mathbf{a}}E_\Delta^\theta(\mathbf{a}), \qquad E_\Delta^\theta(\mathbf{a}) = \frac{1}{2}\mathbf{a}^\top K_\Delta^\theta\mathbf{a}.$$

Here $G_\Delta = G_\Delta^\top \succeq 0$ is the fixed positive-semidefinite structure operator, whereas K_Δ^θ is the trainable diagonal nonnegative matrix.

The Laplace block is a complete-response block: its supervised target is the ambient coefficient response of $\Delta u_{\mathbf{a}}$. For independent training samples $\{\mathbf{a}^{(m)}\}_{m=1}^{N_{\mathrm{tr}}}$, drawn from the coefficient prior above, the relative supervised loss is

$$\mathcal{L}_\Delta(\theta) = \frac{1}{N_{\mathrm{tr}}}\sum_{m=1}^{N_{\mathrm{tr}}}\frac{\left\|\mathbf{F}_\Delta^\theta(\mathbf{a}^{(m)}) - \mathbf{F}_\Delta(\mathbf{a}^{(m)})\right\|_2^2}{\left\|\mathbf{F}_\Delta(\mathbf{a}^{(m)})\right\|_2^2 + \varepsilon_{\mathrm{den}}},$$

where $\varepsilon_{\mathrm{den}} = 10^{-14}$ prevents division by zero. The relative normalization prevents samples with large response amplitudes from dominating the objective. Once trained, the same frozen diffusion block is used for every target domain; only its Galerkin realization changes through N_Ω and M_Ω.

S2.3 Transport blocks

The two directional transport blocks are trained to represent the strong-form mechanisms uu_x and uu_y on the ambient square Q. Because the two mechanisms differ only in the differentiation direction, they share one scalar local density $h_\theta : \mathbb{R} \to \mathbb{R}$. Unlike the coefficient-output supervision used in our previous LegONet implementation[3], the present transfer procedure requires the local primitive derivative itself: the spatial weak derivative is reassembled only after the target geometry has been specified. We therefore supervise the transferable local map $g_\theta(s) = h_\theta'(s)$.

For $\mathbf{a} \in \mathbb{R}^M$, define

$$H_\theta^Q(\mathbf{a}) = \frac{1}{|Q|}\int_Q h_\theta\left(u_{\mathbf{a}}(\mathbf{x})\right)\,d\mathbf{x}.$$

Since the ambient Fourier basis is orthonormal, $\nabla_{\mathbf{a}}H_\theta^Q(\mathbf{a}) = \mathcal{P}_Q[g_\theta(u_{\mathbf{a}})]$. For $\alpha \in \{x, y\}$, let $D_\alpha = \mathcal{P}_Q\partial_\alpha\mathcal{R}_Q$ denote the analytic Fourier differentiation matrix. Both

D_x and D_y are skew-symmetric under the normalized ambient inner product. Taking $J_\alpha = D_\alpha$, we define

$$\mathbf{F}^\theta_\alpha(\mathbf{a}) = J_\alpha \nabla_{\mathbf{a}} H^Q_\theta(\mathbf{a}), \qquad \alpha \in \{x, y\}.$$

Thus $\mathbf{F}^\theta_\alpha$ represents the Galerkin coefficient response of $\partial_\alpha g_\theta(u_{\mathbf{a}})$. The local map g_θ is trained so that this response approximates $u_{\mathbf{a}} \partial_\alpha u_{\mathbf{a}}$. Consequently,

$$\mathbf{F}^\theta_x(\mathbf{a}) \approx \mathcal{P}_Q \left[u_{\mathbf{a}} \partial_x u_{\mathbf{a}}\right], \qquad \mathbf{F}^\theta_y(\mathbf{a}) \approx \mathcal{P}_Q \left[u_{\mathbf{a}} \partial_y u_{\mathbf{a}}\right].$$

The local map g_θ is trained by relative derivative matching:

$$\mathcal{L}_h(\theta) = \frac{\displaystyle\sum_{m=1}^{N_{\mathrm{tr}}} \left| g_\theta(s^{(m)}) - g_{\mathrm{ref}}(s^{(m)}) \right|^2}{\displaystyle\sum_{m=1}^{N_{\mathrm{tr}}} \left| g_{\mathrm{ref}}(s^{(m)}) \right|^2 + \varepsilon_{\mathrm{den}}}.$$

Here $g_{\mathrm{ref}}(s^{(m)})$ is the local reference target associated with $s^{(m)}$, computed offline using the trusted high-order implementation of the transport mechanism. The scalar samples are drawn independently as $s^{(m)} \overset{\mathrm{iid}}{\sim} \mathrm{Uniform}[-1, 1]$. The interval $[-1, 1]$ is the local-amplitude range. Problems with a different amplitude scale can be normalized before applying the transport block, or the scalar training interval can be enlarged. The same trained map g_θ is reused for both spatial directions.

This choice of supervision is essential for transfer. The quantity $g_\theta = h'_\theta$ is local and independent of the target boundary, whereas the derivative, boundary flux, quadrature weights and Galerkin projection are all geometry dependent. Therefore the weak transport response is assembled only after C, N_Ω, M_Ω and the target quadrature have been constructed. In this way, transport follows the same online interface as diffusion through $\mathbf{q}^\Omega_\alpha$, while using a supervised target that is better matched to cross-domain reuse.

S2.4 Training protocol

The diffusion block is trained on coefficient vectors drawn from the Gaussian prior above. The transport density is trained independently on scalar samples. The network $h_\theta : \mathbb{R} \to \mathbb{R}$ is a fully connected multilayer perceptron with four hidden layers, width 128, GELU activations and a scalar output. Its derivative h'_θ is evaluated by automatic differentiation.

For each objective, we use 2×10^4 independent training samples and 4×10^3 independent held-out samples. Optimization uses Adam with learning rate 5×10^{-4} and mini-batch size 16. At any fixed Fourier cutoff K, the trained parameters are frozen and reused in every target-domain experiment. Only the boundary-adapted coordinates, target quadrature and reduced Galerkin assembly are rebuilt.

S2.5 Resolution-matched K-sensitivity protocol

For each Fourier cutoff K used in a resolution study, we train a separate resolution-matched ambient library. For any fixed K, that library is trained once on Q and reused across all target domains without geometry-specific neural retraining.

Separate libraries are necessary because K changes $\mathcal{V}_K(Q)$, the coefficient dimension M, the analytic derivative matrices and the boundary-adapted space constructed on each target domain. Thus changing K changes the ambient discretization itself, not merely a neural-network hyperparameter. Resolution-matched pretraining avoids interpolation, truncation and dimension-mismatch errors between distinct Fourier spaces.

Panels (a,b) of Fig. 2 report held-out errors for the transport primitive derivative and the coefficient-space Laplace response. Panels (c,d) of Fig. 2 report the same diagnostics as K is increased. These results test block accuracy and resolution sensitivity before any target-domain transfer is applied. These are one-off offline costs: after training, each resolution-matched library is reused across all target geometries at that K.

S2.6 Boundary transfer to embedded domains

Let $\Omega \subset Q$ be an embedded target domain with volume quadrature $\{(\mathbf{x}_q, w_q)\}_{q=1}^{N_\mathrm{q}}$. Define $(\Phi_\Omega)_{qj} = \phi_j(\mathbf{x}_q)$, $W_\Omega = \mathrm{diag}(w_1, \ldots, w_{N_\mathrm{q}})$, and $M_\Omega = \Phi_\Omega^\top W_\Omega \Phi_\Omega$. Sampling the target trace condition gives

$$C\mathbf{a}(t) = \mathbf{d}(t).$$

The SVD and mass normalization described in the main text yield $N_\Omega \in \mathbb{R}^{M\times r}$, satisfying $CN_\Omega \approx \mathbf{0}$. Every admissible state is represented as

$$\mathbf{a}(t) = \mathbf{a}_\mathrm{bc}(t) + N_\Omega \mathbf{z}(t),$$

where $C\mathbf{a}_\mathrm{bc}(t) \approx \mathbf{d}(t)$. Let $\Psi_\Omega = \Phi_\Omega N_\Omega \in \mathbb{R}^{N_\mathrm{q}\times r}$ be the boundary-adapted basis evaluated at the target quadrature points. Its reduced mass matrix is $M_r = \Psi_\Omega^\top W_\Omega \Psi_\Omega = N_\Omega^\top M_\Omega N_\Omega$. By construction, $M_r = I_r$.

For diffusion, the frozen ambient response at an admissible state is

$$\mathbf{F}_\Delta^\theta\left(\mathbf{a}_\mathrm{bc}(t) + N_\Omega \mathbf{z}\right) \in \mathbb{R}^M.$$

Its reconstruction on Ω is projected onto the boundary-adapted basis through the Galerkin condition

$$M_r \mathbf{q}_\Delta^\Omega(\mathbf{z}, t) = \Psi_\Omega^\top W_\Omega \Phi_\Omega \mathbf{F}_\Delta^\theta\left(\mathbf{a}_\mathrm{bc}(t) + N_\Omega \mathbf{z}\right).$$

Since $M_r = I_r$, the reduced diffusion response is

$$\mathbf{q}_\Delta^\Omega(\mathbf{z}, t) = N_\Omega^\top M_\Omega \mathbf{F}_\Delta^\theta\left(\mathbf{a}_\mathrm{bc}(t) + N_\Omega \mathbf{z}\right) \in \mathbb{R}^r.$$

Thus no target-domain diffusion network is trained; only the deterministic Galerkin realization changes.

For nonlinear transport, define the nodal state $\mathbf{u}_{\mathrm{q}}(\mathbf{z},t) = \Phi_\Omega(\mathbf{a}_{\mathrm{bc}}(t) + N_\Omega \mathbf{z}) \in \mathbb{R}^{N_{\mathrm{q}}}$, and define $\partial_\alpha \Psi_\Omega = (\partial_\alpha \Phi_\Omega) N_\Omega$ for $\alpha \in \{x, y\}$, where $(\partial_\alpha \Phi_\Omega)_{qj} = \partial_\alpha \phi_j(\mathbf{x}_q)$. Let ψ_j denote the jth boundary-adapted basis function. Integration by parts gives

$$\begin{aligned}(\psi_j, \partial_\alpha g_\theta(u))_\Omega = &- (\partial_\alpha \psi_j, g_\theta(u))_\Omega \\ &+ \int_{\partial\Omega} \psi_j g_\theta(u) n_\alpha \, ds,\end{aligned}$$

where n_α is the α-component of the outward unit normal $\mathbf{n} = (n_x, n_y)$. The corresponding reduced response satisfies

$$M_r \mathbf{q}_\alpha^\Omega(\mathbf{z},t) = -(\partial_\alpha \Psi_\Omega)^\top W_\Omega g_\theta\left(\mathbf{u}_{\mathrm{q}}(\mathbf{z},t)\right) + \mathbf{b}_{\alpha,\partial\Omega}(\mathbf{z},t),$$

where $g_\theta(\mathbf{u}_{\mathrm{q}})$ is applied componentwise and $\mathbf{b}_{\alpha,\partial\Omega}(\mathbf{z},t) \in \mathbb{R}^r$ collects the weak boundary-flux contributions. Its jth entry is

$$\left[\mathbf{b}_{\alpha,\partial\Omega}(\mathbf{z},t)\right]_j = \int_{\partial\Omega} \psi_j(\mathbf{x}) g_\theta\left(u_{\mathbf{a}}(\mathbf{x},t)\right) n_\alpha \, ds, \qquad \mathbf{a}(t) = \mathbf{a}_{\mathrm{bc}}(t) + N_\Omega \mathbf{z}(t).$$

For the homogeneous Dirichlet transport experiments, the tangent basis has numerically vanishing boundary trace, so $\mathbf{b}_{\alpha,\partial\Omega} \approx \mathbf{0}$. Using $M_r = I_r$, the reduced directional responses are therefore

$$\mathbf{q}_\alpha^\Omega(\mathbf{z},t) = -(\partial_\alpha \Psi_\Omega)^\top W_\Omega g_\theta\left(\mathbf{u}_{\mathrm{q}}(\mathbf{z},t)\right), \qquad \alpha \in \{x, y\}.$$

The local reassembly is required because nonlinear projection on Q does not commute with restriction to a new embedded domain. The learned quantity that transfers is $g_\theta = h_\theta'$; the target derivative, boundary flux, quadrature and Galerkin projection are assembled only after N_Ω has been constructed.

S2.7 Physical-to-embedded rescaling.

All benchmark equations are first specified on a physical domain Ω_{phys}, with physical coordinate $\boldsymbol{\xi}$, and mapped to an embedded domain $\Omega \subset Q$, with coordinate $\mathbf{x}$, by

$$T_\Omega : \Omega_{\mathrm{phys}} \to \Omega, \qquad \mathbf{x} = T_\Omega(\boldsymbol{\xi}) = A_\Omega \boldsymbol{\xi} + \mathbf{b}_\Omega, \qquad u(\mathbf{x},t) = u_{\mathrm{phys}}\left(T_\Omega^{-1}(\mathbf{x}), t\right).$$

Here A_Ω is nonsingular and T_Ω uses only translations, rotations and scalings. Thus the affine transformation introduces only constant metric factors in the differential operators; these factors are absorbed into the embedded-domain coefficients. It is therefore a coordinate change, not an additional numerical approximation. Approximation and rollout errors on Ω transfer back to Ω_{phys} up to fixed geometry-dependent norm-equivalence constants.

S3 Manufactured-solution benchmark details

This section gives the analytic domains, manufactured fields, forcing terms, error metrics, quantitative comparisons, and full rollout visualizations for the five manufactured-solution tests used in Fig. 3.

For each manufactured benchmark, we first prescribe an exact field u_{ex} and then define the interior forcing as the residual obtained after substituting u_{ex} into the target differential operator. The boundary data are defined in the same way by applying the prescribed boundary operator $\mathcal{B}$ to u_{ex} on $\partial\Omega$.

The boundary-adapted reduced spaces used by Geometry-aware LegONet are constructed as in the Methods section of the main manuscript; the boundary sample counts and final reduced dimensions are summarized in Fig. 3 of the main text. All trajectories are advanced by fourth-order Runge–Kutta.

S3.1 Error and boundary metrics

Let U_{ex} denote the manufactured solution and U_{num} the numerical approximation evaluated at the same saved times. For a scalar field, the reported relative error is

$$E_{\mathrm{rel}}(t) = \frac{\|u_{\mathrm{num}}(t) - u_{\mathrm{ex}}(t)\|_{L^2(\Omega)}}{\|u_{\mathrm{ex}}(t)\|_{L^2(\Omega)}}.$$

For a vector field $U = (u, v)$, the same definition is used with

$$\|U\|^2_{L^2(\Omega)} = \int_\Omega (u^2 + v^2)\, d\mathbf{x}.$$

Table 1 reports the final, mean, and maximum values of $E_{\mathrm{rel}}(t)$ over the saved evaluation times.

Let $\mathcal{B}$ denote the boundary operator associated with the prescribed boundary condition: $\mathcal{B}U = U$ for Dirichlet data, $\mathcal{B}U = \mathbf{n}\cdot\nabla U$ for Neumann data, and $\mathcal{B}U = \mathbf{n}\cdot\nabla U + \kappa U$ for Robin data, where $\mathbf{n}$ is the outward unit normal. For mixed boundary conditions, the corresponding operator is applied on each boundary component. The exact boundary data are denoted by $g_\partial(\mathbf{x}, t) = \mathcal{B}U_{\mathrm{ex}}(\mathbf{x}, t)$, $\mathbf{x} \in \partial\Omega$. Given boundary diagnostic samples $\{\mathbf{x}_j^b\}_{j=1}^{N_b}$ and saved evaluation times $\{t_\ell\}_{\ell=1}^{N_t}$, the RMS boundary residual is defined as

$$E_\partial^{\mathrm{rms}} = \left(\frac{1}{N_t N_b} \sum_{\ell=1}^{N_t} \sum_{j=1}^{N_b} \left\| \mathcal{B}U_{\mathrm{num}}(\mathbf{x}_j^b, t_\ell) - g_\partial(\mathbf{x}_j^b, t_\ell) \right\|_2^2 \right)^{1/2}.$$

We also report the solution-scaled boundary residual

$$E_\partial^{\mathrm{sc}} = \frac{E_\partial^{\mathrm{rms}}}{S_U + \varepsilon_{\mathrm{sc}}},$$

where

$$S_U = \left(\frac{1}{N_t N_{\mathrm{q}}} \sum_{\ell=1}^{N_t} \sum_{i=1}^{N_{\mathrm{q}}} \|U_{\mathrm{ex}}(\mathbf{x}_i, t_\ell)\|_2^2 \right)^{1/2}$$

is the RMS magnitude of the exact field over the embedded quadrature nodes and saved times. We use $\varepsilon_{\mathrm{sc}} = 10^{-14}$ only to avoid division by zero.

S3.2 MMS-I: rosette Dirichlet reaction–diffusion

The first test is a cubic reaction–diffusion equation with homogeneous Dirichlet data:

$$\begin{cases} u_t = \varepsilon^2 \Delta u + \mu u - \beta u^3 + f^{\Omega}, & \mathbf{x} \in \Omega_{\mathrm{ros}}, \\ u = 0, & \mathbf{x} \in \partial\Omega_{\mathrm{ros}}. \end{cases}$$

Let $z = x + \mathrm{i}y$. The rosette domain is

$$\Omega_{\mathrm{ros}} = \{\mathbf{x} = (x, y) \in [-1, 1]^2 : \; q_{\mathrm{ros}}(x, y) \geq 0\}, \qquad q_{\mathrm{ros}} = 0.53^6 - (x^2 + y^2)^3 + 0.72 \; \mathrm{Re}(z^6),$$

with $\mathrm{Re}(z^6) = x^6 - 15x^4y^2 + 15x^2y^4 - y^6$. The manufactured solution is

$$u_{\mathrm{ex}}(\mathbf{x}, t) = q_{\mathrm{ros}}(x, y) w_{\mathrm{ros}}(x, y, t),$$

so that $u_{\mathrm{ex}} = 0$ on $\partial\Omega_{\mathrm{ros}}$.

The oscillatory factor is defined with $\tau_{\mathrm{ros}} = \omega t$ with $\omega = 12$, $A = 0.42$, and $e(t) = 1.05 + 0.35 \sin(0.40\tau_{\mathrm{ros}}) + 0.25 \cos(1.25\tau_{\mathrm{ros}})$. We set

$$\begin{aligned} w_{\mathrm{ros}} = Ae(t) \Big[& (0.45 + 0.40 \sin(0.65\tau_{\mathrm{ros}}))m_1 + (0.30 + 0.45 \cos(1.10\tau_{\mathrm{ros}} + 0.35))m_2 \\ & + 0.34 \sin(0.95\tau_{\mathrm{ros}})m_3 + 0.22 \cos(1.55\tau_{\mathrm{ros}})m_4 \Big], \end{aligned}$$

where

$$\begin{aligned} m_1 &= \sin(2\pi x + \tau_{\mathrm{ros}}) \cos(\pi y - 0.45\tau_{\mathrm{ros}}), \\ m_2 &= \cos(\pi x - 0.75\tau_{\mathrm{ros}}) \sin(3\pi y + 0.90\tau_{\mathrm{ros}}), \\ m_3 &= \sin(2\pi(x + 0.40y) - 0.55\tau_{\mathrm{ros}}) \sin(\pi(x - 2.10y) + 0.70\tau_{\mathrm{ros}}), \\ m_4 &= \cos(3\pi x + 0.35\tau_{\mathrm{ros}}) \cos(2\pi y - 1.20\tau_{\mathrm{ros}}). \end{aligned}$$

We use $\varepsilon^2 = 0.015$, $\mu = 0.5$, $\beta = 1$, $T = 1.00$, and $\Delta t = 1.5 \times 10^{-3}$.

S3.3 MMS-II: crescent Neumann reaction–diffusion

The second test uses the same cubic reaction–diffusion model with homogeneous Neumann data:

$$\begin{cases} u_t = \varepsilon^2 \Delta u + \mu u - \beta u^3 + f^{\Omega}, & \mathbf{x} \in \Omega_{\mathrm{cre}}, \\ \partial_{\mathbf{n}} u = 0, & \mathbf{x} \in \partial\Omega_{\mathrm{cre}}. \end{cases}$$

The crescent is the difference of two disks,

$$\Omega_{\mathrm{cre}} = \{\mathbf{x} = (x, y) :\ x^2 + y^2 \le R_o^2, \quad (x - d_c)^2 + y^2 \ge R_i^2\},$$

with $R_o = 0.76$, $R_i = 0.60$, and $d_c = 0.34$. Define $q_{\mathrm{cre}} = (R_o^2 - x^2 - y^2)((x - d_c)^2 + y^2 - R_i^2)$. Since $q_{\mathrm{cre}} = 0$ on both boundary components, any field of the form $\tanh(\alpha_s q_{\mathrm{cre}}^2 \mathcal{O})$ satisfies $\partial_{\mathbf{n}} u = 0$ on $\partial\Omega_{\mathrm{cre}}$. The exact field is

$$u_{\mathrm{ex}}(\mathbf{x}, t) = \tanh\big(\alpha_s q_{\mathrm{cre}}(x, y)^2 \mathcal{O}(x, y, t)\big), \qquad \alpha_s = 180.$$

The order parameter combines decaying fine scales, growing coarse modes, and moving droplets. With $\tau_{\mathrm{cre}} = \omega t$, $\omega = 7.5$, $A = 1.0$, $\gamma = 2.0$, $w_f(t) = \exp(-0.65\gamma t)$, and $w_c(t) = 1 - \exp(-\gamma t)$, we use

$$\mathcal{O} = A\Big(0.95 w_f F_{\mathrm{fine}} + 1.30 w_c F_{\mathrm{coarse}} + 0.90 w_c F_{\mathrm{drop}}\Big).$$

The fine component is $F_{\mathrm{fine}} = 0.52 f_1 + 0.34 f_2 + 0.18 f_3$, with shifted coordinates $x' = x - \delta_x(t)$, $y' = y - \delta_y(t)$, where $\delta_x(t) = 0.08 \sin(0.55\tau_{\mathrm{cre}}) - 0.05 \cos(1.10\tau_{\mathrm{cre}})$ and $\delta_y(t) = -0.07 \cos(0.45\tau_{\mathrm{cre}}) + 0.05 \sin(1.35\tau_{\mathrm{cre}})$. The fine modes are

$$\begin{aligned}
f_1 &= \sin(4.8\pi x' + 0.65\tau_{\mathrm{cre}}) \sin(4.2\pi y' - 0.70\tau_{\mathrm{cre}}),\\
f_2 &= \cos(5.4\pi(x' + 0.25y') - 0.45\tau_{\mathrm{cre}}) \sin(3.8\pi(y' - 0.20x') + 0.80\tau_{\mathrm{cre}}),\\
f_3 &= \sin(6.2\pi(x' - 0.10y') + 0.35\tau_{\mathrm{cre}}).
\end{aligned}$$

The coarse component is

$$F_{\mathrm{coarse}} = 0.92 c_1 - 0.78 c_2 + 0.48 c_3,$$

where

$$\begin{aligned}
c_1 &= \sin(1.45\pi(x + 0.32y) + 0.30\tau_{\mathrm{cre}}),\\
c_2 &= \cos(1.15\pi(y - 0.18x) - 0.42\tau_{\mathrm{cre}}),\\
c_3 &= \sin(0.95\pi(0.8x - 1.1y) + 0.55\tau_{\mathrm{cre}}).
\end{aligned}$$

The droplet component is

$$F_{\mathrm{drop}} = 1.15 b_1 - 0.95 b_2 + 0.80 b_3,$$

where b_j are localized Gaussian blobs with time-dependent centers. We use $\varepsilon^2 = 0.01$, $\mu = \beta = 1$, $T = 1.00$, and $\Delta t = 2 \times 10^{-3}$.

S3.4 MMS-III: bunny Robin logistic reaction–diffusion

The third test is a logistic reaction–diffusion equation with nonhomogeneous Robin data:

$$\begin{cases} u_t = D\Delta u + r_g u\left(1 - \dfrac{u}{K_c}\right) + f^{\Omega}, & \mathbf{x} \in \Omega_{\mathrm{bun}}, \\ \partial_{\mathbf{n}} u + \kappa u = g_R, & \mathbf{x} \in \partial\Omega_{\mathrm{bun}}. \end{cases}$$

The bunny boundary is parameterized by an angular boundary parameter $\vartheta \in [0, 2\pi)$: $x_b(\vartheta) = \rho_{\mathrm{bun}}(\vartheta)\cos\vartheta$ and $y_b(\vartheta) = -0.10 + \rho_{\mathrm{bun}}(\vartheta)\sin\vartheta$, where

$$\begin{aligned}\rho_{\mathrm{bun}}(\vartheta) = 0.475 &+ 0.305 G_{\pi/2-0.43,0.255}(\vartheta) + 0.305 G_{\pi/2+0.43,0.255}(\vartheta) - 0.070 G_{\pi/2,0.165}(\vartheta) \\ &+ 0.040 G_{\pi-0.28,0.36}(\vartheta) + 0.040 G_{0.28,0.36}(\vartheta) - 0.025 G_{3\pi/2,0.42}(\vartheta).\end{aligned}$$

Here $G_{\alpha,\omega}(\vartheta) = \exp[-d_{2\pi}(\vartheta,\alpha)^2/(2\omega^2)]$, with $d_{2\pi}$ denoting the periodic angular distance. The outward unit normal $\mathbf{n} = (n_x, n_y)$ is computed from the boundary tangent.

Let $S(s) = (1 + \exp(-s))^{-1}$ be the logistic sigmoid and let $\varphi(t) = c_f t$. The manufactured field is prescribed directly in physical space:

$$u_{\mathrm{ex}}(\mathbf{x}, t) = K_c\, S(2.2\, R(x, y, t))\,.$$

The profile R is defined by

$$R = -1.10 + 1.80F + 1.15C_1 + 0.90C_2 - 0.55H + 0.45W + 0.35 A_{\mathrm{rip}} R_{\mathrm{rip}}.$$

The moving front is $F = S\big(\alpha_s[-0.32 + 0.78(1 - e^{-1.10t}) - (x + 0.10y)]\big)$. The two expanding colonies are $C_1 = S\big(1.15\alpha_s[0.18 + 0.18(1 - e^{-1.6t}) - d_1]\big)$, and $C_2 = S\big(1.05\alpha_s[0.12 + 0.16(1 - e^{-2.0t}) - d_2]\big)$, where $d_1 = [((x - c_{1x})/0.42)^2 + ((y - c_{1y})/0.34)^2]^{1/2}$ and $d_2 = [((x - c_{2x})/0.36)^2 + ((y - c_{2y})/0.28)^2]^{1/2}$. The moving centers are $c_{1x} = -0.34 + 0.05\sin(0.45\varphi)$, $c_{1y} = -0.10 + 0.07\cos(0.60\varphi)$, $c_{2x} = 0.06 + 0.06\cos(0.55\varphi + 0.3)$, and $c_{2y} = 0.24 + 0.05\sin(0.90\varphi)$. The static cavity is $H = S(1.10\alpha_s[0.18 - d_H])$, with $d_H = [((x - 0.28)/0.34)^2 + ((y + 0.22)/0.28)^2]^{1/2}$. The decaying wake and ripple field are $W = 0.12e^{-0.9t}\cos(2.6\pi y - 0.45\varphi)$, and $R_{\mathrm{rip}} = e^{-1.6t}\,[\sin(4\pi x + 0.9\varphi)\cos(3\pi y - 0.6\varphi) + 0.55\cos(2.5\pi(x - y) + 0.4\varphi)]$.

The target parameters are $D = 0.025$, $r_g = 0.8$, $K_c = 1$, $\kappa = 2$, $c_f = 3.2$, $\alpha_s = 3.0$, $A_{\mathrm{rip}} = 0.05$, $T = 0.90$, and $\Delta t = 7.5 \times 10^{-4}$. The nonhomogeneous Robin data are handled by an affine coefficient lift: at each time t, the sampled trace data $g_R = \partial_{\mathbf{n}} u_{\mathrm{ex}} + \kappa u_{\mathrm{ex}}$ are mapped by a right inverse of the Robin trace operator to $\mathbf{a}_{\mathrm{bc}}(t)$, and the remaining dynamics are evolved in the homogeneous Robin-compatible coordinates

$$\mathbf{a}(t) = \mathbf{a}_{\mathrm{bc}}(t) + N_\Omega \mathbf{z}(t).$$

S3.5 MMS-IV: annular-star coupled reaction–diffusion

The fourth test is a two-component coupled reaction–diffusion system:

$$\begin{cases} u_t = D_u \Delta u + a - u + u^2 v + f_u^{\Omega}, & \mathbf{x} \in \Omega_{\mathrm{ann}}, \\ v_t = D_v \Delta v + b - u^2 v + f_v^{\Omega}, & \mathbf{x} \in \Omega_{\mathrm{ann}}, \\ u = v = 0, & \mathbf{x} \in \Gamma_{\mathrm{out}}, \\ \partial_{\mathbf{n}} u = \partial_{\mathbf{n}} v = 0, & \mathbf{x} \in \Gamma_{\mathrm{in}}. \end{cases}$$

Using polar coordinates $\rho = \sqrt{x^2 + y^2}$ and $\theta = \operatorname{atan2}(y, x)$, the annular star is defined by $R_{\mathrm{in}}(\theta) \le \rho \le R_{\mathrm{out}}(\theta)$, where $R_{\mathrm{out}} = 0.70 + 0.105\cos(5\theta) + 0.040\sin(2\theta)$ and $R_{\mathrm{in}} = 0.23 + 0.055\cos(3\theta + \pi/5)$. Let $q_{\mathrm{out}} = R_{\mathrm{out}}(\theta)^2 - \rho^2$, $q_{\mathrm{in}} = \rho^2 - R_{\mathrm{in}}(\theta)^2$, and $s = q_{\mathrm{out}} q_{\mathrm{in}}^2$. The exact fields are

$$u_{\mathrm{ex}} = A_u s \mathcal{U}, \qquad v_{\mathrm{ex}} = A_v s \mathcal{V},$$

with $A_u = 4.0$ and $A_v = 3.2$. Because $s = q_{\mathrm{out}} q_{\mathrm{in}}^2$, the fields satisfy homogeneous Dirichlet data on Γ_{out} and homogeneous Neumann data on Γ_{in}.

The profiles $\mathcal{U}$ and $\mathcal{V}$ are defined from radial bands, rotating angular spots, and oscillatory bridge terms. Let $\eta = q_{\mathrm{in}}/(q_{\mathrm{in}} + q_{\mathrm{out}} + 10^{-12})$. We use the radial bands $B_{\mathrm{in}} = \exp[-((\eta - 0.26)/0.12)^2]$, $B_{\mathrm{mid}} = \exp[-((\eta - 0.50)/0.15)^2]$, and $B_{\mathrm{out}} = \exp[-((\eta - 0.73)/0.13)^2]$. With $\omega = 4.6$, define the angular phases $\varphi_{\mathrm{m}}(t) = \omega t + 0.40\sin(1.25t)$ and $\varphi_{\mathrm{a}}(t) = -0.60\omega t + 0.55\cos(0.95t)$. The angular spot functions are $S_{\mathrm{m}} = \exp[3.1(\cos(\theta - \varphi_{\mathrm{m}}) - 1)]$, $S_{\mathrm{pair}} = \exp[2.8(\cos(2(\theta - \varphi_{\mathrm{a}}) - 1.1) - 1)]$, and $S_{\mathrm{in}} = \exp[2.9(\cos(\theta + 0.45\omega t + 0.8) - 1)]$. We also use the stripe $S_{\mathrm{stripe}} = \cos(4\theta - 1.1t + 0.35\sin(0.7t))$, the bridge oscillation $B_{\mathrm{br}} = \sin(\pi(\eta - 0.18) - 1.05t)$, and the activation factor $\chi(t) = \tanh(3(t - 0.42))$. The two manufactured profiles are

$$\begin{aligned} \mathcal{U} &= 1.35(0.65 + 0.35\sin^2(1.15t))B_{\mathrm{mid}} S_{\mathrm{m}} + 0.95(0.55 + 0.45\cos(0.85t))B_{\mathrm{out}} S_{\mathrm{pair}} \\ &\quad - 0.80\chi(t) B_{\mathrm{in}} S_{\mathrm{in}} + 0.32 B_{\mathrm{mid}} S_{\mathrm{stripe}} B_{\mathrm{br}}, \\ \mathcal{V} &= (0.60 + 0.40\cos(1.05t))B_{\mathrm{mid}} \exp[2.7(\cos(\theta - \varphi_{\mathrm{m}} - 0.85) - 1)] - 0.72 B_{\mathrm{out}} S_{\mathrm{pair}} \\ &\quad + 0.58(0.55 + 0.45\sin^2(0.9t + 0.6))B_{\mathrm{in}}\cos(3\theta + 0.55t) + 0.30\sin(2\pi\eta - 0.85t)\cos(2\theta - 0.45t). \end{aligned}$$

We use $D_u = 0.02$, $D_v = 0.012$, $a = 0.8$, $b = 0.6$, $T = 0.90$, and $\Delta t = 1.5 \times 10^{-3}$.

S3.6 MMS-V: pinwheel-shell Burgers transport

The fifth manufactured test is a scalar Burgers transport problem with homogeneous Dirichlet data:

$$\begin{cases} u_t = \nu \Delta u - u u_x - u u_y + f^{\Omega}, & \mathbf{x} \in \Omega_{\mathrm{pin}}, \\ u = 0, & \mathbf{x} \in \partial\Omega_{\mathrm{pin}}. \end{cases}$$

Let $z = x + \mathrm{i}y$ and $\rho^2 = x^2 + y^2$. The pinwheel-shell domain is

$$\Omega_{\mathrm{pin}} = \{\mathbf{x} = (x, y) \in [-1, 1]^2 : \; q_{\mathrm{pin}}(x, y) \ge 0\},$$

where

$$q_{\mathrm{pin}} = 0.60^8 - \rho^8 + 0.30\,\mathrm{Re}(z^6) - 0.10\,\mathrm{Im}(z^5) + 0.12\rho^2\,\mathrm{Re}(z^3).$$

Here $\mathrm{Re}(z^3) = x^3 - 3xy^2$, $\mathrm{Im}(z^5) = 5x^4y - 10x^2y^3 + y^5$, and $\mathrm{Re}(z^6) = x^6 - 15x^4y^2 + 15x^2y^4 - y^6$. Since $q_{\mathrm{pin}} = 0$ on $\partial\Omega_{\mathrm{pin}}$, multiplying the analytic field by q_{pin} enforces the homogeneous Dirichlet condition exactly.

We define

$$u_{\mathrm{ex}}(\mathbf{x}, t) = 0.82\,u_{\mathrm{low}}(x, y, t) + \big(1 + 0.08\sin(0.31\omega t)\big)u_{\mathrm{osc}}(x, y, t).$$

The low-frequency backbone has the form

$$u_{\mathrm{low}}(x, y, t) = AE(t)\frac{q_{\mathrm{pin}}(x, y)}{0.45}\mathcal{P}_{\mathrm{low}}(x, y, t),$$

where $\mathcal{P}_{\mathrm{low}}$ contains the smooth moving fronts, anisotropic Gaussian blobs, and transported wake used in the pinwheel construction. The off-manifold oscillatory component is

$$u_{\mathrm{osc}} = A\frac{q_{\mathrm{pin}}}{0.45}\left(0.022\,\mathcal{E}_p\sin(21\pi\xi_1)\sin(17\pi\xi_2) + 0.013\,\mathcal{S} + 0.010\,\mathcal{W}_s\right).$$

Here $\xi_1 = 0.74x - 0.36y - 0.14\sin(0.43\omega t)$ and $\xi_2 = 0.29x + 0.81y + 0.11\cos(0.37\omega t)$. The packet envelope is $\mathcal{E}_p = \exp[-((x - x_p)/0.44)^2 - ((y - y_p)/0.34)^2]$, with $x_p = -0.06 + 0.18\cos(0.41\omega t + 0.5)$ and $y_p = 0.03 - 0.16\sin(0.46\omega t - 0.2)$. The sheared wave is

$$\mathcal{S} = \sin(19\pi(0.63x + 0.28y) - 0.95\omega t)\exp\left[-\left(\frac{0.52x - 0.61y + 0.10\cos(0.33\omega t)}{0.28}\right)^2\right],$$

and the localized swirl wave is

$$\mathcal{W}_s = \cos(17\pi(x + 0.22y) + 0.61\omega t)\exp\left[-\left(\frac{x - x_s}{0.24}\right)^2 - \left(\frac{y - y_s}{0.24}\right)^2\right],$$

with $x_s = 0.14\sin(0.57\omega t + 0.8)$ and $y_s = -0.17\cos(0.49\omega t - 0.4)$.

The reported target parameters are $\nu = 0.035$, $A = 2.6$, $\omega = 4.0$, $T = 0.90$, and $\Delta t = 1.2 \times 10^{-3}$.

S3.7 Full quantitative comparison and rollout visualizations

Table 1 reports the quantitative comparison on the five MMS tests. Final, mean, and max errors are relative $L^2(\Omega)$ errors on the target MMS case. The last two columns report the RMS boundary-operator residual and its solution-scaled value. For PINN, FNO, and UNO, the time column reports target training time. For Geometry-aware LegONet, no target-domain neural block is trained; the reported time is the sum of target-domain SVD/null-space construction, reduced-operator assembly, and rollout time. Fig. S1 shows the corresponding multi-time rollouts.

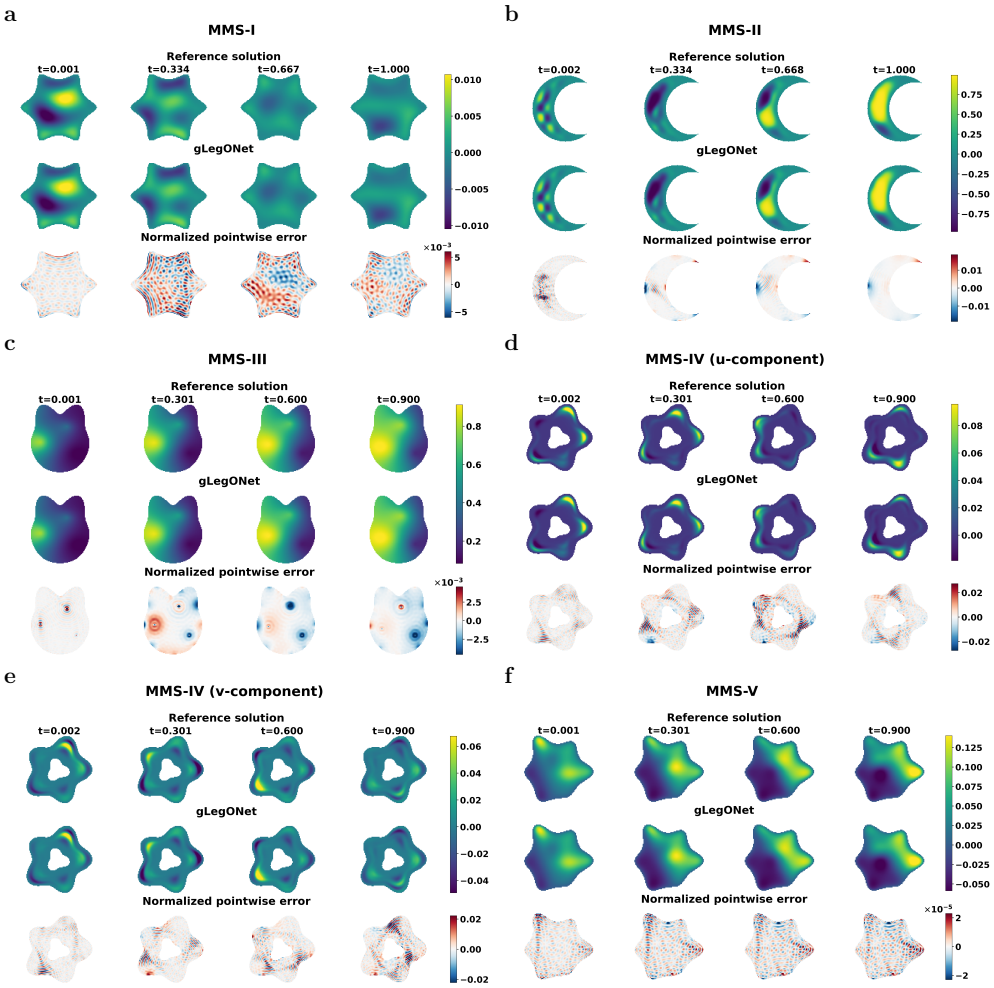

Fig. S1: Transferred-block rollouts on the manufactured-solution benchmarks. **a**, MMS-I rosette reaction–diffusion. **b**, MMS-II crescent reaction–diffusion. **c**, MMS-III bunny Robin logistic dynamics. **d**, MMS-IV u-component on the annular-star domain. **e**, MMS-IV v-component on the annular-star domain. **f**, MMS-V pinwheel-shell Burgers transport. Each panel shows the reference solution, Geometry-aware LegONet prediction and normalized pointwise error at representative rollout times.

S4 Baseline configurations

This section describes the baseline solvers used in the manufactured-solution, volume-constrained Allen–Cahn, and inner-square Burgers experiments. The PINN baseline is a label-free per-instance coordinate solver. The FNO and UNO baselines are supervised grid-based neural time-steppers. They are trained from analytic manufactured trajectories, finite-element trajectories, or finite-volume trajectories, depending on

the experiment. For the inner-square Burgers experiment, we also include a direct geometry-specific block baseline constructed on the target square, following our previous work[3].

Let $\Omega \subset [-1,1]^2$ denote the target domain and let $\mathbf{U}$ denote the state variable, either a scalar field u or a vector field (u,v). The problem data consist of the embedded geometry, initial condition, forcing or reference generator, boundary operator, boundary data, and PDE parameters. All reported errors are evaluated on the same target trajectory and with the same metrics as in the corresponding main-text experiment.

S4.1 Fourier-feature PINN

The PINN baseline is a per-instance coordinate solver trained separately for each target PDE and domain. The network maps $(\mathbf{x},t)$ to the physical field, with one output channel for scalar problems and two output channels for coupled or vector problems. We use a Fourier-feature MLP with five hidden layers, width 128, and tanh activations. The input is first mapped to

$$\Gamma(\mathbf{x},t) = \left[x, y, \tau, \{\sin(2^k\pi x), \cos(2^k\pi x), \sin(2^k\pi y), \cos(2^k\pi y), \sin(2^k\pi\tau), \cos(2^k\pi\tau)\}_{k=0}^{K_f-1}\right],$$

with $\tau = t/T$, $K_f = 6$. The initial condition is imposed by the hard-constrained ansatz

$$\widehat{\mathbf{U}}_\eta(\mathbf{x},t) = \mathbf{U}_0(\mathbf{x}) + \frac{t}{T} s_U \mathcal{N}_\eta(\mathbf{x},t),$$

where $\mathcal{N}_\eta$ is the trainable Fourier-feature neural network and $s_U > 0$ is an output scale estimated from the target problem. The prefactor t/T forces the network correction to vanish at $t = 0$, and therefore $\widehat{\mathbf{U}}_\eta(\cdot,0) = \mathbf{U}_0$ holds exactly.

Interior collocation points $\{(\mathbf{x}_i,t_i)\}_{i=1}^{N_r}$ are sampled by drawing spatial points uniformly from the embedding box and rejecting those outside Ω; time points are sampled uniformly from the current training interval. Boundary points $\{(\mathbf{x}_j^\partial,t_j^\partial)\}_{j=1}^{N_\partial}$ are sampled from $\partial\Omega$ and the same time interval. The PINN loss is a strong-form residual loss plus a boundary penalty:

$$\mathcal{L}_{\mathrm{PINN}} = \frac{1}{N_r}\sum_{i=1}^{N_r} \left\| \mathcal{R}_{\mathrm{PDE}}\left(\widehat{\mathbf{U}}_\eta;\mathbf{f}\right)(\mathbf{x}_i,t_i)\right\|_2^2 + w_b\mathcal{L}_{\partial\Omega}.$$

Here $\mathcal{R}_{\mathrm{PDE}}$ is the strong-form residual of the target equation, $\mathbf{f}$ denotes the corresponding forcing or source field when present. The boundary loss $\mathcal{L}_{\partial\Omega}$ is selected according to the target boundary operator:

$$\mathcal{L}_{\partial\Omega} = \frac{1}{N_\partial}\sum_{j=1}^{N_\partial} \left\| \mathcal{B}\widehat{\mathbf{U}}_\eta(\mathbf{x}_j^\partial,t_j^\partial) - \mathbf{g}_\partial(\mathbf{x}_j^\partial,t_j^\partial)\right\|_2^2.$$

Thus $\mathcal{B}\widehat{\mathbf{U}}_\eta - \mathbf{g}_\partial$ equals $\widehat{\mathbf{U}}_\eta - \mathbf{g}_D$ for Dirichlet data, $\mathbf{n}\cdot\nabla\widehat{\mathbf{U}}_\eta - \mathbf{g}_N$ for Neumann data, and $\mathbf{n}\cdot\nabla\widehat{\mathbf{U}}_\eta + \kappa\widehat{\mathbf{U}}_\eta - \mathbf{g}_R$ for Robin data, where $\mathbf{n}$ is the outward unit normal. For

mixed boundary conditions, the corresponding penalty is applied on each boundary component.

Unless otherwise stated, we use $N_r = 4096$, $N_\partial = 1024$, $w_b = 5$, and 2.0×10^4 AdamW iterations. The learning rate is 10^{-3}, the weight decay is 10^{-8}, the gradient clipping threshold is 1, and a cosine schedule decays the learning rate to 5% of its initial value. A time curriculum expands the collocation window to the full interval $[0, T]$ over the first 1.5×10^4 iterations.

S4.2 FNO and UNO autoregressive baselines

The FNO and UNO baselines are implemented as autoregressive neural time-steppers. Let $\mathbf{U}^n$ denote the channel-stacked physical state at the saved time t_n, and let $\mathbf{U}_0$ be the prescribed initial condition. The evaluation rollout is initialized by $\widehat{\mathbf{U}}^{\,0} = \mathbf{U}_0$, so the initial condition is imposed strongly.

We normalize the state channelwise. With training-set statistics $\boldsymbol{\mu}_U$ and $\boldsymbol{\sigma}_U$, define

$$\mathbf{W}^n = \frac{\mathbf{U}^n - \boldsymbol{\mu}_U}{\boldsymbol{\sigma}_U}, \qquad \widehat{\mathbf{W}}^{\,n} = \frac{\widehat{\mathbf{U}}^{\,n} - \boldsymbol{\mu}_U}{\boldsymbol{\sigma}_U}.$$

All divisions in this definition are componentwise. In particular,

$$\widehat{\mathbf{W}}^{\,0} = \frac{\mathbf{U}_0 - \boldsymbol{\mu}_U}{\boldsymbol{\sigma}_U}.$$

At step n, the neural-operator input is assembled as

$$\mathbf{X}^n = \big[\widehat{\mathbf{W}}^{\,n}, \mathcal{A}^n\big],$$

where $\widehat{\mathbf{W}}^{\,n}$ is the current predicted normalized state and $\mathcal{A}^n$ denotes the auxiliary channels available for the experiment, such as the embedded mask, coordinate channels, normalized time, forcing channels, and parameter channels. The exact contents of $\mathcal{A}^n$ are specified in the experiment-specific baseline subsections.

Let

$$\mathcal{G}_\eta : \mathbf{X}^n \mapsto \Delta\mathbf{W}^n_\eta$$

denote the trainable FNO or UNO increment model with parameters η. The network predicts an increment in normalized variables,

$$\Delta\mathbf{W}^n_\eta = \mathcal{G}_\eta(\mathbf{X}^n),$$

and the normalized state is advanced by the residual update

$$\widehat{\mathbf{W}}^{\,n+1} = \widehat{\mathbf{W}}^{\,n} + \Delta\mathbf{W}^n_\eta.$$

The corresponding physical prediction is obtained by inverse normalization,

$$\widehat{\mathbf{U}}^{\,n+1} = \boldsymbol{\sigma}_U \odot \widehat{\mathbf{W}}^{\,n+1} + \boldsymbol{\mu}_U,$$

where $\odot$ denotes componentwise multiplication. For masked-grid experiments, the domain mask is applied to the physical output after each update.

Training uses a multi-step autoregressive rollout loss with horizon $H = 5$. For a minibatch of B reference sequences $\{\mathbf{U}_m^{n+j}\}_{j=0}^{H}$, where $m = 1, \ldots, B$, the rollout is initialized from $\mathbf{U}_m^n$ and unrolled for H neural steps. The data loss is

$$\mathcal{L}_{\mathrm{roll}} = \frac{1}{BH} \sum_{m=1}^{B} \sum_{j=1}^{H} \frac{\left\| \chi_\Omega \left(\widehat{\mathbf{U}}_m^{n+j} - \mathbf{U}_m^{n+j} \right) \right\|_2^2}{\left\| \chi_\Omega \mathbf{U}_m^{n+j} \right\|_2^2 + \varepsilon_{\mathrm{roll}}}.$$

Here B is the minibatch size, χ_Ω is the binary embedded-domain mask on the Cartesian training grid, applied componentwise to all state channels, and $\varepsilon_{\mathrm{roll}} = 10^{-14}$ is used only to avoid division by zero. The full neural-operator objective is

$$\mathcal{L}_{\mathrm{op}} = \mathcal{L}_{\mathrm{roll}} + \lambda_{\mathrm{bc}} \mathcal{L}_{\partial\Omega}.$$

The boundary loss $\mathcal{L}_{\partial\Omega}$ uses the residual associated with the target boundary operator. For Dirichlet data, predicted grid fields are bilinearly interpolated to the sampled boundary points and compared with the prescribed trace. For Neumann and Robin data, spatial gradients are first computed on the Cartesian grid using finite-difference stencils, then interpolated to the sampled boundary points and contracted with the outward normal. The parameter λ_{bc} weights this soft boundary penalty. Unless otherwise stated, we use $\lambda_{\mathrm{bc}} = 1$.

The FNO baseline uses a pointwise lifting layer, four Fourier layers, and a pointwise projection. Unless otherwise stated, we use a 96^2 grid, channel width 48, and 16 Fourier modes in each spatial direction. The UNO baseline follows the U-shaped neural-operator design of Rahman et al.[4], with the same grid, width, and 12 retained Fourier modes. Both models are trained with 5000 optimizer steps, AdamW, minibatch size 4, weight decay 10^{-6}, gradient clipping threshold 1, and a cosine learning-rate schedule whose final learning rate is 5% of its initial value. The FNO learning rate is 2×10^{-3}, and the UNO learning rate is 7.5×10^{-4}. The best checkpoint is selected by the held-out multi-step validation loss.

S4.3 Manufactured-solution baseline protocol

This subsection describes the FNO and UNO baselines used for the MMS tests. Reference sequences are generated directly from analytic manufactured-solution families. For the qth MMS benchmark, let $\boldsymbol{\vartheta} \in \Theta^{(q)}$ denote the low-dimensional parameter vector indexing the manufactured-solution family, such as amplitudes, frequencies, or front-speed parameters. The target values, training ranges, and target-exclusion convention are specified in Table S1. For each training sample, we draw

$$\boldsymbol{\vartheta} \sim \mathrm{Uniform}\left(\Theta_{\mathrm{train}}^{(q)} \right),$$

Table S1: Parameterized manufactured families used to train the supervised FNO and UNO baselines. Target parameters are the held-out values used in the five MMS benchmarks; training ranges define the surrounding amortized trajectory families.

Test	Target parameters	Training ranges
MMS-I	$A = 0.42,\ \omega = 12$	$A \in [0.30, 0.55],\ \omega \in [9, 15]$
MMS-II	$A = 1.0,\ \omega = 7.5$	$A \in [0.80, 1.20],\ \omega \in [6.50, 8.50]$
MMS-III	$c_f = 3.2,\ \alpha_s = 3.0,\ A_{\mathrm{rip}} = 0.05$	$c_f \in [2.70, 3.70],\ \alpha_s \in [2.40, 3.60],\ A_{\mathrm{rip}} \in [0.02, 0.08]$
MMS-IV	$A_u = 4.0,\ A_v = 3.2,\ \omega = 4.6$	$A_u \in [3.40, 4.60],\ A_v \in [2.70, 3.70],\ \omega \in [3.80, 5.40]$
MMS-V	$A = 2.6,\ \omega = 4.0$	$A \in [2.00, 3.20],\ \omega \in [3.20, 4.80]$

while rejecting samples in the 12% exclusion window around the target parameter $\boldsymbol{\vartheta}^{(q)}_{\mathrm{target}}$ specified in that table. This exclusion prevents the supervised FNO and UNO baselines from training directly on the target MMS case.

For a given parameter vector $\boldsymbol{\vartheta}$, let $\mathbf{f}^{\Omega}_{\boldsymbol{\vartheta}}(\mathbf{x}, t)$ denote the forcing or source field obtained by substituting the corresponding manufactured solution into the target PDE. At saved time level n, the MMS neural-operator input follows the common autoregressive form

$$\mathbf{X}^n = \big[\widehat{\mathbf{W}}^{\,n}, \mathcal{A}^n_{\mathrm{MMS}}\big],$$

where $\widehat{\mathbf{W}}^{\,n}$ is the current predicted normalized state and

$$\mathcal{A}^n_{\mathrm{MMS}} = \big[\chi_\Omega, x, y, t_n/T, \widetilde{\mathbf{f}}^{\,n}, \widetilde{\mathbf{f}}^{\,n+1}, \widetilde{\boldsymbol{\vartheta}}\big].$$

The auxiliary tensor contains the embedded mask, coordinate channels, normalized time, normalized forcing channels at the current and next saved time levels, and the normalized family-parameter vector. Here $\mathbf{f}^n = \mathbf{f}^{\Omega}_{\boldsymbol{\vartheta}}(\cdot, t_n)$ and $\mathbf{f}^{n+1} = \mathbf{f}^{\Omega}_{\boldsymbol{\vartheta}}(\cdot, t_{n+1})$, and $\mathbf{f}^n$, $\mathbf{f}^{n+1}$, and $\boldsymbol{\vartheta}$ are normalized using the corresponding training-set statistics.

S4.4 Volume-constrained Allen–Cahn disk baseline protocol

This subsection describes the FNO and UNO baselines used for the Allen–Cahn equation. This test has no analytic target trajectory, so the supervised neural-operator baselines are trained from independently generated finite-element trajectories. The evaluation reference is a P_1 finite-element solution on a disk mesh with 900 boundary nodes and 20000 interior nodes. The PINN baseline follows the common PINN protocol in Section S4.1.

The FNO and UNO baselines are autoregressive neural time-steppers on a masked 96×96 Cartesian grid. At saved time level n, the input tensor is

$$\mathbf{X}^n = \big[\widehat{\mathbf{W}}^{\,n}, \mathcal{A}_{\mathrm{AC}}\big], \qquad \mathcal{A}_{\mathrm{AC}} = \big[\chi_{\Omega_{\mathrm{disk}}}, x, y\big].$$

Here $\widehat{\mathbf{W}}^{\,n}$ is the current predicted normalized state, defined as in Section S4.2, $\chi_{\Omega_{\mathrm{disk}}}$ is the binary disk mask on the Cartesian grid, and x, y are coordinate channels.

Training trajectories are generated by the same finite-element solver used for the evaluation reference, but from initial conditions disjoint from the target trajectory. The

target and baseline initial conditions are sampled from the same smooth reduced-space interface family on the disk. This family constructs mean-zero, Neumann-compatible interfacial fields in the boundary-adapted trial space and adds only low-frequency seeded perturbations, rather than arbitrary grid noise.

In the reported runs, the target trajectory uses seed $s_0 = 1234$, interface amplitude 0.70, nominal bending parameter 0.28, and smoothness parameter $m_{\mathrm{ic}} = 18$. The supervised FNO/UNO dataset uses 18 disjoint seeds: the first 8 generated trajectories are used for training, the next 2 for validation, and the remaining 8 for held-out testing. Each initial condition is projected to the finite-element space, evolved to $T_{\mathrm{train}} = 4$ with solver step $\Delta t_{\mathrm{FEM}} = 5 \times 10^{-4}$, and sampled every $\Delta T = 0.05$ to form autoregressive training sequences. Homogeneous Neumann data are imposed through the soft boundary penalty with $N_{\mathrm{bc}}^{\mathrm{train}} = 2048$ boundary samples.

S4.5 Inner-square vector Burgers baseline protocol

This subsection describes the baselines used for the Burgers equation. The target problem is a two-component viscous Burgers equation on the aligned square $\Omega = [-h, h]^2$, with $h = 0.55$, homogeneous Dirichlet data, $\nu = 10^{-2}$, and $T = 0.5$. We write the physical state as $\mathbf{U} = (u, v)^\top$.

The PINN baseline is a per-instance coordinate solver following Section S4.1, with two output channels. Its strong-form residual contains the two Burgers equations. The initial condition is imposed by the hard-constrained PINN ansatz, and the homogeneous Dirichlet boundary data are imposed by a soft penalty on $|u_\eta|^2 + |v_\eta|^2$ over boundary samples.

The FNO and UNO baselines are supervised autoregressive time-steppers trained from finite-volume trajectories. The dataset consists of Dirichlet-compatible random sine-series initial conditions. Reference trajectories are generated on a uniform finite-volume grid with solver step $\Delta t_{\mathrm{FV}} = 5 \times 10^{-4}$, final time $T = 0.5$, and snapshots saved every 20 solver steps. Hence the neural time-step size is $\Delta T = 10^{-2}$. The default dataset contains 64 training trajectories, 8 validation trajectories, and 8 test trajectories.

At step n, the neural-operator input is

$$\mathbf{X}^n = \big[\widehat{\mathbf{W}}^n, \mathcal{A}_{\mathrm{Burg}}\big], \qquad \mathcal{A}_{\mathrm{Burg}} = \big[x/h, y/h\big],$$

where $\widehat{\mathbf{W}}^n$ is the current predicted normalized two-component state, and $x/h, y/h$ are normalized coordinate channels on the square grid. No mask channel is needed because the computational grid coincides with the square domain. The FNO and UNO architectures and optimizers follow the common protocol in Section S4.2.

The direct inner-square block baseline follows the geometry-specific block construction used in our previous work[3], in particular the two-dimensional vector Burgers experiment reported as Experiment A4 in the Supplementary Information of that work. Unlike the Geometry-aware LegONet ambient-transfer solver, this baseline does not use blocks pretrained on the ambient square $Q = [-1, 1]^2$. Instead, the basis and operator blocks are constructed directly on $\Omega = [-h, h]^2$, using basis functions that satisfy the homogeneous Dirichlet condition.

The direct solver uses the tensor-product sine basis

$$\psi_{mn}(x,y) = h^{-1}\sin\left(\frac{m\pi(x+h)}{2h}\right)\sin\left(\frac{n\pi(y+h)}{2h}\right), \qquad m,n \geq 1.$$

The two velocity components are represented as

$$u(x,y,t) = \sum_{m,n} c^u_{mn}(t)\psi_{mn}(x,y), \qquad v(x,y,t) = \sum_{m,n} c^v_{mn}(t)\psi_{mn}(x,y).$$

This basis is orthonormal in $L^2(\Omega)$ and satisfies $\psi_{mn} = 0$ on $\partial\Omega$. Therefore both velocity components are represented in a boundary-compatible coefficient space, and $u = v = 0$ is enforced to roundoff by construction. The time step, final time, nonlinear substepping, initial condition, retained modal dimension, and finite-volume reference grid are matched to the ambient-transfer experiment. This comparison isolates the effect of boundary transfer: the direct block is specialized to the inner square, whereas Geometry-aware LegONet transfers a fixed ambient block library through the boundary-adapted coefficient space without target-domain neural retraining.

S5 Downstream dynamics: problem definitions and diagnostics

This section provides the complete numerical configurations for the four downstream tests reported in the main text: volume-constrained Allen–Cahn dynamics on a disk, vector Burgers dynamics on an embedded square, vorticity transport past a cylinder, and clamped Swift–Hohenberg dynamics on a heart-shaped domain. For each test, we specify the target PDE, boundary-adapted realization, time integrator, reference solver, diagnostic metrics and ablation protocol.

S5.1 Volume-constrained Allen–Cahn dynamics on a disk

The target domain is $\Omega_{\text{disk}} = \{\mathbf{x} \in Q : x^2 + y^2 \leq 0.4^2\}$. We solve

$$\begin{cases} \partial_t u = \varepsilon^2 \Delta u + u - u^3 - \lambda(t), & \mathbf{x} \in \Omega_{\text{disk}}, \quad 0 < t \leq T, \\ \partial_{\mathbf{n}} u = 0, & \mathbf{x} \in \partial\Omega_{\text{disk}}, \end{cases}$$

where

$$\lambda(t) = \frac{1}{|\Omega_{\text{disk}}|}\int_{\Omega_{\text{disk}}} \left(u - u^3\right)\, d\mathbf{x}.$$

Integrating the PDE over Ω_{disk} and using the homogeneous Neumann condition gives $\frac{d}{dt}\int_{\Omega_{\text{disk}}} u\, d\mathbf{x} = 0$; the spatial mean is therefore conserved by the continuous problem. We use $\varepsilon^2 = 5 \times 10^{-3}$, $T = 4$, and $\Delta t = 5 \times 10^{-4}$.

Because the boundary data are homogeneous, $\mathbf{a}_{\text{bc}}(t) = \mathbf{0}$ and $\mathbf{a}(t) = N_\Omega \mathbf{z}(t)$. The Neumann constraint is constructed from $N_b = 1600$ boundary samples using

$\tau_C = 10^{-10}$, and mass normalization gives the retained dimension $r = 412$. Boundary residuals are evaluated on an independent set of $N_b^{\mathrm{val}} = 6400$ boundary points.

Let $\Psi_\Omega = \Phi_\Omega N_\Omega$ and $\mathbf{u}_{\mathrm{q}}(\mathbf{z}) = \Psi_\Omega \mathbf{z} \in \mathbb{R}^{N_{\mathrm{q}}}$. The discrete mean correction is

$$\lambda_h(\mathbf{z}) = \frac{\mathbf{1}_{\mathrm{q}}^\top W_\Omega \left[\mathbf{u}_{\mathrm{q}}(\mathbf{z}) - \mathbf{u}_{\mathrm{q}}(\mathbf{z})^{\odot 3}\right]}{\mathbf{1}_{\mathrm{q}}^\top W_\Omega \mathbf{1}_{\mathrm{q}}},$$

where $\mathbf{1}_{\mathrm{q}} = (1, \ldots, 1)^\top \in \mathbb{R}^{N_{\mathrm{q}}}$ is the all-ones vector on the volume quadrature grid, and $\odot$ denotes componentwise multiplication. The numerator approximates $\int_\Omega (u - u^3)\, d\mathbf{x}$, while the denominator approximates $|\Omega|$. The reduced dynamics advanced by the transferred solver are

$$\dot{\mathbf{z}} = \varepsilon^2 \mathbf{q}_\Delta^\Omega(\mathbf{z}) + \Psi_\Omega^\top W_\Omega \left[\mathbf{u}_{\mathrm{q}}(\mathbf{z}) - \mathbf{u}_{\mathrm{q}}(\mathbf{z})^{\odot 3} - \lambda_h(\mathbf{z}) \mathbf{1}_{\mathrm{q}}\right].$$

Here $\mathbf{q}_\Delta^\Omega$ is the target-domain response of the frozen Laplace block. The reaction and mean-correction terms are evaluated at the volume quadrature points and projected through the same boundary-adapted Galerkin interface.

Time integration uses symmetric Strang splitting. Each reaction half-step is advanced by fourth-order Runge–Kutta with two substeps, and the diffusion step is advanced by Crank–Nicolson. The mean correction λ_h is recomputed at every nonlinear Runge–Kutta stage.

An independent P_1 finite-element solution is used as reference. The mesh contains 900 boundary nodes and 20000 interior nodes. The same continuous initial field is projected independently into the finite-element space and the boundary-adapted Fourier space. All field errors are evaluated on a common embedded quadrature grid.

For the numerical solution, define

$$\bar{u}_h(t) = \frac{\mathbf{1}_{\mathrm{q}}^\top W_\Omega \mathbf{u}_{\mathrm{q}}(t)}{\mathbf{1}_{\mathrm{q}}^\top W_\Omega \mathbf{1}_{\mathrm{q}}}, \qquad D_{\mathrm{mass}}^{\max} = \max_{t \in \mathcal{T}_{\mathrm{eval}}} \left|\bar{u}_h(t) - \bar{u}_h(0)\right|.$$

The transferred solver obtains final, mean and maximum relative $L^2(\Omega_{\mathrm{disk}})$ errors of 2.98×10^{-2}, 3.18×10^{-2}, and 5.83×10^{-2}, respectively. Its dense Neumann residual is 3.47×10^{-9}, and $D_{\mathrm{mass}}^{\max} = 3.63 \times 10^{-5}$ (Table 2). The rollout, relative-error history and spatial mean are shown in Fig. 4a–c.

The rank sweep in Fig. 4e shows that $r = 80$, 120, 200, 300, and 412 give final errors 7.50×10^{-1}, 6.34×10^{-1}, 1.15×10^{-1}, 3.37×10^{-2}, and 2.98×10^{-2}, respectively. By contrast, the boundary-sampling sweep in panel (d) changes the error only weakly once a few hundred Neumann samples are used, while the residual on the independent validation set remains at the 10^{-9} level. For this smooth domain, reduced-space capacity is therefore the dominant limitation after the sampled trace has been resolved.

S5.2 Vector Burgers dynamics on an embedded inner square

The target domain is $\Omega = [-h, h]^2 \subset Q$, with $h = 0.55$. We solve

$$\begin{cases} u_t + uu_x + vu_y = \nu \Delta u, & (x, y) \in \Omega, \quad 0 < t \le T, \\ v_t + uv_x + vv_y = \nu \Delta v, & (x, y) \in \Omega, \quad 0 < t \le T, \\ u = v = 0, & (x, y) \in \partial\Omega, \end{cases}$$

with $\nu = 10^{-2}$ and $T = 0.5$. The initial conditions are

$$u_0(\mathbf{x}) = \sin\left(\frac{\pi(x+h)}{2h}\right) \sin\left(\frac{\pi(y+h)}{2h}\right),$$
$$v_0(\mathbf{x}) = \sin\left(\frac{\pi(x+h)}{h}\right) \sin\left(\frac{\pi(y+h)}{2h}\right),$$

and both components vanish on $\partial\Omega$.

The same boundary-adapted matrix N_Ω is used for both velocity components. It is constructed from $N_b = 1000$ Dirichlet boundary samples, has retained dimension $r = 833$, and is validated on 4200 independent boundary points. We write

$$\mathbf{a}_u = N_\Omega \mathbf{z}_u, \qquad \mathbf{a}_v = N_\Omega \mathbf{z}_v, \qquad \Psi_\Omega = \Phi_\Omega N_\Omega.$$

The self-advection terms are realized by the frozen directional transport blocks. For the cross-advection terms, define

$$\mathbf{c}_\alpha^\Omega(\mathbf{z}_m, \mathbf{z}_q) = \Psi_\Omega^\top W_\Omega \left[(\Psi_\Omega \mathbf{z}_m) \odot ((\partial_\alpha \Psi_\Omega)\mathbf{z}_q) \right], \qquad \alpha \in \{x, y\}.$$

The transferred reduced system is

$$\dot{\mathbf{z}}_u = \nu \mathbf{q}_\Delta^\Omega(\mathbf{z}_u) - \mathbf{q}_x^\Omega(\mathbf{z}_u) - \mathbf{c}_y^\Omega(\mathbf{z}_v, \mathbf{z}_u),$$
$$\dot{\mathbf{z}}_v = \nu \mathbf{q}_\Delta^\Omega(\mathbf{z}_v) - \mathbf{c}_x^\Omega(\mathbf{z}_u, \mathbf{z}_v) - \mathbf{q}_y^\Omega(\mathbf{z}_v).$$

Here $\mathbf{q}_x^\Omega$ and $\mathbf{q}_y^\Omega$ are the target-domain weak realizations of the pretrained $u\,u_x$ and $u\,u_y$ mechanisms.

Time integration uses symmetric Strang splitting. Nonlinear half-steps are advanced by RK4 with six substeps, and the diffusion step is advanced by Crank–Nicolson. The outer time step is $\Delta t = 5 \times 10^{-4}$. The reference is an independently implemented 201×201 upwind finite-volume solver with the same physical time step.

In addition to relative field errors, we evaluate consistency with the homogeneous-boundary kinetic-energy identity. Define

$$E(t) = \frac{1}{2} \int_\Omega \left(u^2 + v^2\right) d\mathbf{x}.$$

For an exact solution satisfying $u = v = 0$ on $\partial\Omega$,

$$\frac{dE}{dt} = \frac{1}{2}\int_\Omega \left(u^2 + v^2\right)\left(\partial_x u + \partial_y v\right)\, d\mathbf{x} - \nu \int_\Omega \left(|\nabla u|^2 + |\nabla v|^2\right)\, d\mathbf{x}.$$

For each method m, let $U^{\mathsf{m}}(t) = (u^{\mathsf{m}}(t), v^{\mathsf{m}}(t))$ denote its numerical rollout evaluated on the common diagnostic quadrature grid. We compute the normalized kinetic-budget residual

$$R_E^{\mathsf{m}}(t) = \frac{\left|\dot{E}^{\mathsf{m}}(t) - \mathcal{P}^{\mathsf{m}}(t) + \mathcal{D}^{\mathsf{m}}(t)\right|}{|\dot{E}^{\mathsf{m}}(t)| + |\mathcal{P}^{\mathsf{m}}(t)| + |\mathcal{D}^{\mathsf{m}}(t)| + 10^{-14}},$$

where $\dot{E}^{\mathsf{m}}(t)$ is the discrete time derivative of the kinetic energy computed from U^{m},

$$\mathcal{P}^{\mathsf{m}}(t) = \frac{1}{2}\int_\Omega \left((u^{\mathsf{m}})^2 + (v^{\mathsf{m}})^2\right)\left(\partial_x u^{\mathsf{m}} + \partial_y v^{\mathsf{m}}\right) dx$$

is the advective energy-production term, and

$$\mathcal{D}^{\mathsf{m}}(t) = \nu \int_\Omega \left(|\nabla u^{\mathsf{m}}|^2 + |\nabla v^{\mathsf{m}}|^2\right)\, dx$$

is the viscous dissipation. The last column of Table 2 reports $R_E^{\mathrm{final}} = R_E^{\mathsf{m}}(T)$ for each method. For an exact solution of the continuous Burgers system, this residual is zero; hence smaller values indicate better consistency with the kinetic-energy balance.

The transferred solver obtains final, mean and maximum relative errors 1.83×10^{-2}, 1.17×10^{-2}, and 1.83×10^{-2}, respectively. Its dense Dirichlet residual is 1.91×10^{-13}, and its final kinetic-budget residual is 3.63×10^{-7} (Table 2). These values match the geometry-specific direct inner-square block solver to the reported precision, although the latter uses a sine basis constructed specifically for Ω. The componentwise rollouts and error histories are shown in Fig. 5a–c.

Fig. 5d varies the embedded half-width h. Accuracy deteriorates when the target square occupies an increasingly large fraction of the fixed ambient box, identifying the resolution limit of the $K = 22$ library. Panel (e) shows that additional boundary samples have little effect once the Dirichlet trace is resolved, whereas panel (f) shows that increasing r improves the rollout until the dynamically relevant boundary-admissible modes have been retained.

S5.3 Navier–Stokes vorticity transport past an embedded cylinder

Let $\boldsymbol{\xi} = (\xi_1, \xi_2)$ denote the physical coordinate, and let $B_R(\boldsymbol{\xi}_0) = \left\{\boldsymbol{\xi} \in \mathbb{R}^2 : \|\boldsymbol{\xi} - \boldsymbol{\xi}_0\|_2 < R\right\}$ denote the open disk of radius R centered at $\boldsymbol{\xi}_0$. The physical domain is

$$\Omega_{\mathrm{phys}} = [0, 8] \times [-1, 1] \setminus B_{R_{\mathrm{cyl}}}(\boldsymbol{\xi}_{\mathrm{cyl}}), \qquad \boldsymbol{\xi}_{\mathrm{cyl}} = (1.15, 0), \qquad R_{\mathrm{cyl}} = 0.22.$$

It is mapped into Q by the isotropic affine transformation

$$\mathbf{x} = T_\Omega(\boldsymbol{\xi}) = s_\Omega\left(\boldsymbol{\xi} - (4,0)\right), \qquad s_\Omega = 0.2375.$$

The outer rectangle is therefore mapped to $[-0.95, 0.95] \times [-0.2375, 0.2375]$. All differential operators, velocities and diffusion coefficients used by the embedded solver are transformed consistently under T_Ω; the constant metric factors are absorbed into the embedded coefficients.

In physical coordinates, the target system is

$$\begin{cases} \partial_t \omega + \left(\mathbf{u}_{\mathrm{bg}} + \nabla_{\boldsymbol{\xi}}^{\perp}\psi\right) \cdot \nabla_{\boldsymbol{\xi}}\omega = \nu \Delta_{\boldsymbol{\xi}}\omega + f, & \boldsymbol{\xi} \in \Omega_{\mathrm{phys}}, \\ -\Delta_{\boldsymbol{\xi}}\psi = \omega, & \boldsymbol{\xi} \in \Omega_{\mathrm{phys}}, \\ \omega = \psi = 0, & \boldsymbol{\xi} \in \partial\Omega_{\mathrm{phys}}, \end{cases}$$

where $\nabla_{\boldsymbol{\xi}}^{\perp}\psi = (\partial_{\xi_2}\psi, -\partial_{\xi_1}\psi)$. We use $\nu = 10^{-2}$ and $T = 1.8$.

The initial vorticity, background velocity and source terms are multiplied by a boundary cutoff. Let $S(s) = \widetilde{s}^2(3 - 2\widetilde{s})$, $\widetilde{s} = \min\{1, \max\{0, s\}\}$. For $\rho(\boldsymbol{\xi}) = \|\boldsymbol{\xi} - \boldsymbol{\xi}_{\mathrm{cyl}}\|_2$, define $d_{\mathrm{out}}(\boldsymbol{\xi}) = \min\{\xi_1,\, 8 - \xi_1,\, \xi_2 + 1,\, 1 - \xi_2\}$, $d_{\mathrm{cyl}}(\boldsymbol{\xi}) = \rho(\boldsymbol{\xi}) - R_{\mathrm{cyl}}$, $d_\partial(\boldsymbol{\xi}) = \min\{d_{\mathrm{out}}(\boldsymbol{\xi}), d_{\mathrm{cyl}}(\boldsymbol{\xi})\}$, and

$$\chi_\partial(\boldsymbol{\xi}) = \mathbf{1}_{\Omega_{\mathrm{phys}}}(\boldsymbol{\xi}) S\left(\frac{d_\partial(\boldsymbol{\xi})}{0.10}\right).$$

Hence all prescribed fields vanish continuously at both the channel wall and the cylinder.

The background velocity is a regularized, wall-damped cylinder-bypass field. With $X = \xi_1 - 1.15$, $Y = \xi_2$, $\rho^2 = X^2 + Y^2$, and $\rho_{\mathrm{reg}}^4 = \max\{\rho^4, (0.15 R_{\mathrm{cyl}})^4\}$, define

$$\mathbf{u}_{\mathrm{bg}}^{\mathrm{raw}} = U_{\mathrm{bg}} \left(1 - \frac{R_{\mathrm{cyl}}^2 (X^2 - Y^2)}{\rho_{\mathrm{reg}}^4}, -\frac{2 R_{\mathrm{cyl}}^2 XY}{\rho_{\mathrm{reg}}^4}\right), \qquad U_{\mathrm{bg}} = 1.55.$$

The raw field is multiplied by

$$q_{\mathrm{bg}} = S\left(\frac{\rho - R_{\mathrm{cyl}}}{0.12}\right) S\left(\frac{1 - |\xi_2|}{0.12}\right),$$

giving $\mathbf{u}_{\mathrm{bg}} = \mathbf{1}_{\Omega_{\mathrm{phys}}} q_{\mathrm{bg}} \mathbf{u}_{\mathrm{bg}}^{\mathrm{raw}}$.

The initial vorticity is an antisymmetric Gaussian dipole. With $\xi_{1,0} = 1.15 - 0.68$, $d_0 = 0.14$, $\sigma_1 = 0.13$, and $\sigma_2 = 0.06$, define

$$G_\pm(\boldsymbol{\xi}) = \exp\left[-\frac{(\xi_1 - \xi_{1,0})^2}{2\sigma_1^2} - \frac{(\xi_2 \mp d_0/2)^2}{2\sigma_2^2}\right],$$

and $\omega_0(\boldsymbol{\xi}) = A_0 \chi_\partial(\boldsymbol{\xi}) \left(G_+(\boldsymbol{\xi}) - G_-(\boldsymbol{\xi})\right)$, where A_0 is chosen so that $\|\omega_0\|_{L^\infty(\Omega_{\mathrm{phys}})} = 1.5$.

Let $\boldsymbol{\xi}_{\mathrm{cyl}} = (c_1, c_2)$ and R_{cyl} denote the cylinder centre and radius. The source combines an attached shear layer, a finite train of alternating vortex packets and a downstream oscillatory perturbation:

$$f(\boldsymbol{\xi}, t) = \chi_\partial(\boldsymbol{\xi}) \left[f_S(\boldsymbol{\xi}, t) + f_P(\boldsymbol{\xi}, t) + f_C(\boldsymbol{\xi}, t) \right].$$

The cutoff χ_∂ defined above forces the complete source to vanish continuously on both the outer channel wall and the cylinder. For compactness, define the Gaussian envelope $\mathcal{G}_{\sigma_1,\sigma_2}(\boldsymbol{\xi}; \boldsymbol{\mu}) = \exp[-(\xi_1 - \mu_1)^2/(2\sigma_1^2) - (\xi_2 - \mu_2)^2/(2\sigma_2^2)]$, and the temporal ramp $\mathcal{R}(t; T) = 1 - \exp[-(t/T)^2]$. The attached shear source is

$$f_S(\boldsymbol{\xi}, t) = \mathcal{R}(t; T_S)\, A_S\, S\!\left(\frac{\xi_1 - (c_1 + R_{\mathrm{cyl}} + s_0)}{s_w} \right) \mathcal{G}_{\sigma_{S,1},\sigma_{S,2}}\left(\boldsymbol{\xi}; (\xi_{S,1}, c_2)\right) \frac{\xi_2 - c_2}{\sigma_{S,2}}.$$

It is localized immediately behind the cylinder and is odd about $\xi_2 = c_2$, thereby initiating a weak attached shear layer. We use $A_S = 0.12$, $s_0 = 0.05$, $s_w = 0.30$, $\xi_{S,1} = c_1 + R_{\mathrm{cyl}} + 0.34$, $\sigma_{S,1} = 0.55$, $\sigma_{S,2} = 0.24$, and $T_S = 0.18$. The packet source emits $N_{\mathrm{pkt}} = 7$ alternating-sign vortex pairs from $\xi_{\mathrm{emit}} = c_1 + R_{\mathrm{cyl}} + x_P$, with $x_P = 0.20$. For $m = 0, \ldots, N_{\mathrm{pkt}} - 1$, let $\tau_m = mT_P$ and $a_m(t) = t - \tau_m$. Its activation, streamwise centre, vertical displacement and waviness are

$$\begin{aligned}
\Theta_m(t) &= \mathbf{1}_{\{a_m(t) \geq 0\}}\, S\!\left(\frac{a_m(t)}{T_{\mathrm{rise}}} \right) \exp\!\left(-\frac{a_m(t)}{T_{\mathrm{decay}}} \right), \\
\xi_{1,m}(t) &= \xi_{\mathrm{emit}} + V_P a_m(t), \\
\eta_m(t) &= y_P \exp\!\left(-\frac{a_m(t)}{T_{\mathrm{bend}}} \right), \\
w_m(t) &= A_w \sin(\omega_w a_m(t) + m\phi_P).
\end{aligned}$$

Here $\mathbf{1}_{\{\cdot\}}$ is the indicator function. The packet forcing is

$$\begin{aligned}
f_P(\boldsymbol{\xi}, t) = A_P \sum_{m=0}^{N_{\mathrm{pkt}}-1} (-1)^m \Theta_m(t) \Big[& \mathcal{G}_{\sigma_{P,1},\sigma_{P,2}}\left(\boldsymbol{\xi}; \left(\xi_{1,m}(t), c_2 + \eta_m(t) + w_m(t)\right)\right) \\
& - \gamma_P \mathcal{G}_{\sigma_{P,1},\sigma_{P,2}}\left(\boldsymbol{\xi}; \left(\xi_{1,m}(t), c_2 - \eta_m(t) + 0.5 w_m(t)\right)\right) \Big].
\end{aligned}$$

The parameters are $A_P = 2.0$, $T_P = 0.16$, $T_{\mathrm{rise}} = 0.035$, $T_{\mathrm{decay}} = 0.65$, $V_P = 1.65$, $\sigma_{P,1} = 0.20$, $\sigma_{P,2} = 0.105$, $y_P = 0.17$, $\gamma_P = 0.95$, $T_{\mathrm{bend}} = 0.60$, $A_w = 0.055$, $\omega_w = 5.5$, and $\phi_P = 1.4$. Finally, the downstream perturbation is

$$f_C(\boldsymbol{\xi}, t) = \mathcal{R}(t; T_C)\, A_C\, \mathcal{G}_{\sigma_{C,1},\sigma_{C,2}}\left(\boldsymbol{\xi}; (\xi_{C,1}, c_2)\right) \sin(\omega_C t + \phi_C),$$

where $\xi_{C,1} = c_1 + R_{\mathrm{cyl}} + x_{\mathrm{col}}$. We use $A_C = 0.22$, $x_{\mathrm{col}} = 1.00$, $\sigma_{C,1} = 0.48$, $\sigma_{C,2} = 0.24$, $\omega_C = 12.0$, $\phi_C = 0$, and $T_C = 0.20$. The three components play distinct roles:

f_S initiates the attached shear layer, f_P generates alternating vortex packets, and f_C induces a mild downstream interaction. Their common cutoff preserves boundary compatibility while producing a nontrivial wake evolution.

The streamfunction is recovered in the same homogeneous Dirichlet boundary-adapted space as the vorticity. Let $\Psi_\Omega = \Phi_\Omega N_\Omega$, and represent their values at the volume quadrature points by $\boldsymbol{\omega}_{\mathrm{q}} = \Psi_\Omega \mathbf{z}_\omega$ and $\boldsymbol{\psi}_{\mathrm{q}} = \Psi_\Omega \mathbf{z}_\psi$. Since $\Psi_\Omega^\top W_\Omega \Psi_\Omega = I_r$, the weak form of $-\Delta\psi = \omega$ gives

$$K_P \mathbf{z}_\psi = \mathbf{z}_\omega, \qquad K_P = (\partial_x \Psi_\Omega)^\top W_\Omega (\partial_x \Psi_\Omega) + (\partial_y \Psi_\Omega)^\top W_\Omega (\partial_y \Psi_\Omega).$$

The matrix $K_P \in \mathbb{R}^{r\times r}$ is assembled once.

The induced velocity at the quadrature points is

$$\nabla^\perp \psi_h = \left((\partial_y \Psi_\Omega)\mathbf{z}_\psi, -(\partial_x \Psi_\Omega)\mathbf{z}_\psi\right).$$

This velocity is added to the embedded background field and used in the vorticity-transport update.

The reported solver uses $K = 45$, $r = 6099$, a 300×96 volume grid, 6000 boundary samples for constructing N_Ω, and 2×10^4 independent boundary samples for validation. Diffusion is provided by the frozen Laplace response, the streamfunction by the reduced Poisson solve, and transport by an upwind discretization of the total velocity. The rollout uses IMEX Euler with $\Delta t = 5 \times 10^{-4}$. The reference is an independently implemented masked-grid finite-difference solver using implicit diffusion and the same upwind transport convention.

At $K = 45$, the final, mean and maximum relative errors are 3.85×10^{-2}, 4.96×10^{-2}, and 7.44×10^{-2}, respectively, and the dense boundary residual is 1.59×10^{-12}. Increasing K from 22 to 45 reduces the mean error from 1.07×10^{-1} to 4.96×10^{-2} and the maximum error from 1.77×10^{-1} to 7.44×10^{-2}. The final-time error saturates near $K = 40$ (Table 3 and Fig. S2).

S5.4 Clamped Swift–Hohenberg manufactured path on a heart-shaped domain

The embedded domain $\Omega_{\mathrm{heart}} \subset Q = [-1, 1]^2$ is the interior of a heart-shaped curve. Let $\vartheta \in [0, 2\pi)$, $x_0(\vartheta) = \sin^3 \vartheta$, and $y_0(\vartheta) = 0.82 \cos\vartheta - 0.28 \cos(2\vartheta) - 0.08 \cos(3\vartheta) - 0.02 \cos(4\vartheta)$. We use the shape factors $N_{\mathrm{notch}}(\vartheta) = 1 - \eta_{\mathrm{notch}} \exp[-\alpha_x \cos^2 \vartheta - \alpha_y(\sin\vartheta - c_{\mathrm{notch}})^2]$, $S_{\mathrm{side}}(\vartheta) = 1 + \eta_{\mathrm{q}} \cos(2\vartheta)$, and $B_{\mathrm{bend}}(\vartheta) = 1 + \eta_{\mathrm{b}} \sin\vartheta$, with $a = b = 0.62$, $\eta_{\mathrm{notch}} = 0.30$, $\alpha_x = 10$, $\alpha_y = 30$, $c_{\mathrm{notch}} = 0.75$, $\eta_{\mathrm{q}} = 0.08$, $\eta_{\mathrm{b}} = -0.40$, and vertical shift $s_{\mathrm{h}} = -0.05$. The unrotated boundary is

$$\begin{aligned} \widetilde{x}_b(\vartheta) &= a\, S_{\mathrm{side}}(\vartheta) N_{\mathrm{notch}}(\vartheta) x_0(\vartheta), \\ \widetilde{y}_b(\vartheta) &= b\, B_{\mathrm{bend}}(\vartheta) y_0(\vartheta) + s_{\mathrm{h}}. \end{aligned}$$

The final boundary is obtained by rotating $(\widetilde{x}_b, \widetilde{y}_b)$ through $\alpha_{\mathrm{h}} = -0.10$. The outward unit normal $\mathbf{n}_b = (n_{x,b}, n_{y,b})$ is computed from the boundary tangent and oriented away from the curve centroid.

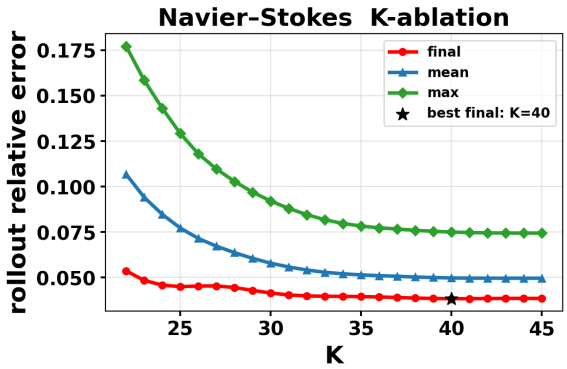

Fig. S2: Resolution ablation for vorticity transport past an embedded cylinder. Increasing K enriches the ambient coefficient space and reduces the mean and maximum relative errors. The final-time error saturates near $K = 40$, showing that boundary enforcement remains stable while the interior rollout is limited by ambient resolution and retained admissible rank.

We solve the clamped Swift–Hohenberg equation

$$\begin{cases} \partial_t u = -\gamma(\Delta + k_0^2 I)^2 u + \mu u - u^3 + f, & \mathbf{x} \in \Omega_{\text{heart}}, \quad 0 < t \le T, \\ u = \partial_{\mathbf{n}} u = 0, & \mathbf{x} \in \partial\Omega_{\text{heart}}, \end{cases}$$

with $\gamma = 4.0 \times 10^{-3}$, $k_0 = 5.3$, $\mu = 0.05$, $T = 4.0$, and $\Delta t = 10^{-3}$.

The clamped boundary-adapted space is built from the ambient trigonometric basis $\{\phi_j\}_{j=1}^{M}$ with $K = 30$, giving $M = 2821$. On $N_b = 1200$ boundary samples, the value and normal-trace matrices are $(C_D)_{b,j} = \phi_j(\mathbf{x}_b)$ and $(C_N)_{b,j} = \mathbf{n}_b^\top \nabla\phi_j(\mathbf{x}_b)$. After row normalization, the clamped constraint matrix is

$$C_{\text{cl}} = \begin{bmatrix} C_D^{\text{norm}} \\ C_N^{\text{norm}} \end{bmatrix}.$$

The null space is computed with relative SVD tolerance $\tau_C = 10^{-10}$. After mass normalization with $\tau_M = 10^{-10}$, the retained dimension is selected by the mass tolerance, giving $r = 795$. On 3600 independent boundary points, the dense value- and normal-trace residuals are 4.38×10^{-10} and 1.79×10^{-8}, respectively.

Let $N_\Omega \in \mathbb{R}^{M \times r}$ be the resulting coefficient matrix, $\Psi_\Omega = \Phi_\Omega N_\Omega$ the volume-quadrature basis, W_Ω the diagonal quadrature-weight matrix, and $M_r = \Psi_\Omega^\top W_\Omega \Psi_\Omega$. The mass normalization gives $M_r \approx I_r$. Let $\mathbf{F}_{\Delta,\theta} : \mathbb{R}^M \to \mathbb{R}^M$ denote the frozen ambient Laplace block trained on Q. No fourth-order block is trained. For the jth reduced coordinate direction $\mathbf{e}_j \in \mathbb{R}^r$, define

$$\mathbf{g}_j^\theta = \mathbf{F}_{\Delta,\theta}(N_\Omega \mathbf{e}_j) + k_0^2 N_\Omega \mathbf{e}_j.$$

The composed fourth-order weak matrix is assembled as

$$(K_{\mathrm{SH}}^{\theta})_{ij} = (\mathbf{g}_i^{\theta})^{\top} M_{\Omega} \mathbf{g}_j^{\theta}.$$

The reduced system is

$$M_r \dot{\mathbf{z}} = -\gamma K_{\mathrm{SH}}^{\theta} \mathbf{z} + \mu M_r \mathbf{z} - \mathbf{q}_{u^3}^{\Omega}(\mathbf{z}) + \mathbf{q}_f^{\Omega}(t),$$

where $\mathbf{q}_{u^3}^{\Omega}(\mathbf{z}) = \Psi_{\Omega}^{\top} W_{\Omega} (\Psi_{\Omega} \mathbf{z})^{\odot 3}$, and $\odot$ denotes componentwise multiplication. Since $M_r \approx I_r$, the implementation uses the mass-normalized form of this system.

The manufactured path is prescribed in physical space. Let $d_h(\mathbf{x}) = \min_{\mathbf{x}_b \in \partial\Omega_{\mathrm{heart}}} \|\mathbf{x} - \mathbf{x}_b\|_2$, $d_{\max} = \max_{\Omega_{\mathrm{heart}}} d_h$, and $\rho(\mathbf{x}) = d_h(\mathbf{x})/d_{\max}$. To keep the field coupled to the boundary while preserving the clamped trace, we use the boundary-interacting envelope

$$E_h(\mathbf{x}) = \frac{\rho(\mathbf{x})^2 \left[0.55 + 1.20 \exp\left(-\left(\frac{\rho(\mathbf{x}) - 0.38}{0.22}\right)^2\right)\right]}{\max_{\Omega_{\mathrm{heart}}} \rho^2 \left[0.55 + 1.20 \exp\left(-\left(\frac{\rho - 0.38}{0.22}\right)^2\right)\right]}.$$

This envelope vanishes quadratically at $\partial\Omega_{\mathrm{heart}}$, but its maximum is shifted away from the domain centre. Let $S(s) = 3s^2 - 2s^3$, $s(t) = S(\min\{t/T_a, 1\})$, $T_a = 2.0$, and $\chi(s) = \exp[-((s - 0.55)/0.32)^2]$.

The spatial pattern is built from rotated anisotropic Gaussian lobes. For a centre (x_c, y_c), widths σ_x, σ_y, and angle θ, set

$$G_{\theta}(x, y; x_c, y_c, \sigma_x, \sigma_y) = \exp\left[-\frac{\xi_{\theta}^2}{2\sigma_x^2} - \frac{\eta_{\theta}^2}{2\sigma_y^2}\right],$$

with

$$\begin{pmatrix} \xi_{\theta} \\ \eta_{\theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x - x_c \\ y - y_c \end{pmatrix}.$$

The primary positive and negative lobe centres are $(x_p, y_p) = (-0.25 + 0.48s,\ 0.24 - 0.23s + 0.04\sin(\pi s))$ and $(x_n, y_n) = (0.25 - 0.45s,\ -0.06 - 0.20\sin(\pi s) + 0.22s)$, with orientations $\theta_p = -0.95 + 1.45s$ and $\theta_n = 0.55 - 1.35s$. The corresponding lobes are $G_{\theta_p}(x, y; x_p, y_p, 0.105 + 0.030\chi, 0.205 + 0.050\chi)$ and $G_{\theta_n}(x, y; x_n, y_n, 0.115 + 0.035\chi, 0.215 + 0.055\chi)$. In addition, a positive boundary ribbon is centred at $(x_{r,+}, y_{r,+}) = (-0.18 + 0.30s,\ 0.30 - 0.08s)$ with widths $(0.150, 0.055)$ and angle $-0.10 + 0.55s$, while a negative boundary ribbon is centred at $(x_{r,-}, y_{r,-}) = (0.18 - 0.25s,\ -0.26 + 0.10s)$ with widths $(0.165, 0.060)$ and angle $0.65 - 0.45s$. We denote the corresponding ribbon Gaussians by

$$G_{\mathrm{red,rib}} = G_{-0.10+0.55s}(x, y; x_{r,+}, y_{r,+}, 0.150, 0.055),$$
$$G_{\mathrm{blue,rib}} = G_{0.65-0.45s}(x, y; x_{r,-}, y_{r,-}, 0.165, 0.060).$$

We also use the low-frequency wave

$$W_h(x,y,s) = [\cos(1.2\pi s)(0.75x - 0.20y) + \sin(1.2\pi s)(0.30x + 0.85y)]\, e^{-0.85(x^2+y^2)}.$$

The unnormalized red and blue families are

$$\begin{aligned} R_h(x,y,t) &= E_h(x,y)\left[0.88\, G_{\theta_p} + 0.48\, G_{\mathrm{red,rib}} + 0.16\, \chi(s) W_h\right], \\ B_h(x,y,t) &= E_h(x,y)\left[0.88\, G_{\theta_n} + 0.55\, G_{\mathrm{blue,rib}} - 0.12\, \chi(s) W_h\right]. \end{aligned}$$

Each family is normalized by its maximum absolute value over the volume quadrature points. With $a_R(s) = 0.98 + 0.12\sin(2\pi s) + 0.10\chi(s)$ and $a_B(s) = 0.96 - 0.10\sin(2\pi s + 0.4) + 0.08\chi(s)$, the manufactured field is defined by

$$u_{\mathrm{ref}}(x,y,t) = A \frac{a_R(s)\widehat{R}_h(x,y,t) - a_B(s)\widehat{B}_h(x,y,t) - 0.03\,\overline{a_R\widehat{R}_h - a_B\widehat{B}_h}}{\max_{\Omega_{\mathrm{heart}}} \left| a_R(s)\widehat{R}_h - a_B(s)\widehat{B}_h - 0.03\,\overline{a_R\widehat{R}_h - a_B\widehat{B}_h} \right|}, \qquad A = 0.26.$$

Here hats denote the normalized red and blue families, and the overline denotes the quadrature mean. Because $s(t) = 1$ for $t \geq T_a$, the reference pattern is steady after the active stage. The forcing $\mathbf{q}_f^{\Omega}(t)$ is obtained from the prescribed manufactured path.

We use an IMEX Crank–Nicolson step, treating the composed fourth-order linear part implicitly and the cubic term explicitly. The final, mean and maximum relative $L^2(\Omega_{\mathrm{heart}})$ errors are 7.01×10^{-3}, 8.33×10^{-3}, and 2.08×10^{-2}, respectively. The Fourier-resolution study is reported in Table 3. Over the tested range $K = 22, \ldots, 30$, the dense clamped-boundary residual remains small, between 3.44×10^{-9} and 1.53×10^{-8}, showing that the simultaneous value and normal traces are retained after increasing the ambient cutoff. The final error decreases from 1.47×10^{-2} at $K = 22$ to 7.01×10^{-3} at $K = 30$, while the maximum error decreases from 4.62×10^{-2} to 2.08×10^{-2}. The mean error follows the same trend, decreasing from 1.76×10^{-2} to 8.33×10^{-3}. Thus the dominant effect of increasing K is improved representation of the boundary-interacting physical path, while the clamped boundary realization remains stable.

S5.5 Boundary-adapted physical-law identification details

This section gives the complete configuration of the inverse-discovery experiments reported in Fig. 7. The target geometry and homogeneous Dirichlet boundary condition are assumed known; the unknowns are the active interior mechanisms and their scalar coefficients. No neural parameter is updated during identification.

All experiments use the frozen $K = 22$ ambient library and a boundary-adapted space of rank $r = 80$. Since the boundary data are homogeneous, $\mathbf{a}_{\mathrm{bc}}(t) = \mathbf{0}$ and $\mathbf{a}(t) = N_\Omega \mathbf{z}(t)$. The target-domain candidate dictionary is

$$\mathcal{D}_\theta^{\Omega} = \left\{ \mathbf{q}_{\Delta}^{\Omega},\, \mathbf{q}_x^{\Omega},\, \mathbf{q}_y^{\Omega},\, \mathbf{q}_u^{\Omega},\, \mathbf{q}_{u^2}^{\Omega},\, \mathbf{q}_{u^3}^{\Omega} \right\}.$$

The first three entries are the realized diffusion and directional quadratic-transport responses. The remaining entries are analytic Galerkin projections of u, u^2, and u^3.

Using the ordering above, the unknown reduced dynamics are written as

$$\dot{\mathbf{z}}(t) \approx \sum_{p=1}^{6} c_p \, \mathbf{q}_p^{\Omega}(\mathbf{z}(t), t),$$

where $\mathbf{c} = (c_1, \dots, c_6)^\top$ is expected to be sparse.

The first target is the peanut-shaped dumbbell

$$\Omega_{\mathrm{p}} = \{(\rho\cos\vartheta, \rho\sin\vartheta) : 0 \le \rho < 0.52\,(1 + 0.42\cos(2\vartheta))\}\,.$$

Its target law is

$$\partial_t u = 0.04\Delta u + 0.06u - 0.35u^3, \qquad u = 0 \quad \text{on } \partial\Omega_{\mathrm{p}}.$$

The corresponding coefficient vector is

$$\mathbf{c}_{\mathrm{p}}^{\star} = (0.04,\, 0,\, 0,\, 0.06,\, 0,\, -0.35)^\top.$$

This geometry tests whether reaction–diffusion can be identified across a narrow bottleneck.

The second target is an oblique sinusoidal channel. In local coordinates (ξ, η), define

$$-0.78 \le \xi \le 0.78, \qquad \left|\eta - 0.12\sin\left(\frac{2\pi\xi}{0.78}\right)\right| \le 0.18.$$

The local domain is rotated through $\pi/6$, so that

$$\mathbf{x} = R_{\pi/6}\begin{pmatrix}\xi\\ \eta\end{pmatrix}, \qquad R_{\varphi} = \begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}.$$

The target law is

$$\partial_t u = 0.02\Delta u - 0.06u\,\partial_x u + 0.09u\,\partial_y u - 0.03u + 0.025u^2 - 0.30u^3, \qquad u = 0 \quad \text{on } \partial\Omega_{\mathrm{ch}}.$$

Its coefficient vector is

$$\mathbf{c}_{\mathrm{ch}}^{\star} = (0.02,\, -0.06,\, 0.09,\, -0.03,\, 0.025,\, -0.30)^\top.$$

This case combines nonlinear transport, reaction terms and a thin anisotropic geometry.

For each random seed, independent smooth initial states are generated directly in the homogeneous boundary-adapted space. Each trajectory is observed at fixed random interior sensors $\{\mathbf{x}_s\}_{s=1}^{N_{\mathrm{obs}}}$, which are reused over all observation times and short trajectories for that seed. The same sensor locations are used by all sparse-data methods in a given comparison. The identification protocol uses $N_{\mathrm{traj}} = 16$

independent short trajectories over $T_{\mathrm{id}} = 0.012$, sampled every $\Delta t_{\mathrm{obs}} = 5 \times 10^{-5}$. We use $N_{\mathrm{obs}} = 1000$ fixed random interior sensors in the noise study and $N_{\mathrm{obs}} = 240$ in the sparse-sensor ablation. The identified law is validated by a held-out rollout to $T_{\mathrm{roll}} = 1.0$, using time step 10^{-3}.

Let $t_n = n\Delta t_{\mathrm{obs}}$, $n = 0, \dots, N_t$, denote the observation times on each short trajectory. For trajectory m at time t_n, the noiseless sensor vector is

$$\mathbf{y}_{m,n} = \left(u^{(m)}(\mathbf{x}_1, t_n), \dots, u^{(m)}(\mathbf{x}_{N_{\mathrm{obs}}}, t_n)\right)^\top \in \mathbb{R}^{N_{\mathrm{obs}}}.$$

Observation noise is added independently as

$$\mathbf{y}^{\sigma}_{m,n} = \mathbf{y}_{m,n} + \sigma s_y \boldsymbol{\epsilon}_{m,n}, \qquad \boldsymbol{\epsilon}_{m,n} \sim \mathcal{N}(\mathbf{0}, I_{N_{\mathrm{obs}}}),$$

where s_y is the observation-amplitude normalization. Let $(\Phi_{\mathrm{obs}})_{sj} = \phi_j(\mathbf{x}_s)$. For homogeneous boundary data, the reduced state at each observation time is recovered by the least-squares sensor fit

$$\widehat{\mathbf{z}}_{m,n} = \arg\min_{\mathbf{z} \in \mathbb{R}^r} \left\| \Phi_{\mathrm{obs}} N_\Omega \mathbf{z} - \mathbf{y}^{\sigma}_{m,n} \right\|_2^2.$$

Temporal derivatives $\dot{\widehat{\mathbf{z}}}_{m,n}$ are then estimated along each recovered short trajectory by finite differences.

In the peanut and channel law-identification experiments, the boundary data are static and homogeneous. Hence Eq. (7) reduces to fitting the recovered time derivatives by a linear combination of candidate reduced responses. Let $\mathcal{I}_{\mathrm{der}} \subset \{0, \dots, N_t\}$ denote the time indices at which the finite-difference derivative is used, and define

$$\mathcal{S}_{\mathrm{st}} = \{(m,n) : m = 1, \dots, N_{\mathrm{traj}},\ n \in \mathcal{I}_{\mathrm{der}}\}, \qquad N_{\mathrm{st}} = |\mathcal{S}_{\mathrm{st}}|.$$

For each stacked sample $(m,n) \in \mathcal{S}_{\mathrm{st}}$, define

$$\Theta_{m,n} = \left[\mathbf{q}_1^\Omega(\widehat{\mathbf{z}}_{m,n}, t_n) \ \cdots \ \mathbf{q}_6^\Omega(\widehat{\mathbf{z}}_{m,n}, t_n)\right] \in \mathbb{R}^{r \times 6}.$$

Stacking all derivative-valid samples from all short trajectories gives

$$\Theta = \begin{bmatrix} \Theta_{m_1,n_1} \\ \vdots \\ \Theta_{m_{N_{\mathrm{st}}},n_{N_{\mathrm{st}}}} \end{bmatrix} \in \mathbb{R}^{N_{\mathrm{st}} r \times 6}, \qquad \dot{\widehat{\mathbf{z}}}_{\mathrm{st}} = \begin{bmatrix} \dot{\widehat{\mathbf{z}}}_{m_1,n_1} \\ \vdots \\ \dot{\widehat{\mathbf{z}}}_{m_{N_{\mathrm{st}}},n_{N_{\mathrm{st}}}} \end{bmatrix} \in \mathbb{R}^{N_{\mathrm{st}} r}.$$

The ridge regression problem used in the reported experiments is

$$\widehat{\mathbf{c}} = \arg\min_{\mathbf{c} \in \mathbb{R}^6} \left\| \Theta \mathbf{c} - \dot{\widehat{\mathbf{z}}}_{\mathrm{st}} \right\|_2^2 + \lambda_{\mathrm{ridge}} \|\mathbf{c}\|_2^2.$$

When the active support is unknown, we use sequential thresholded ridge least squares in the normalized dictionary. Starting from all six candidates, we solve the ridge problem on the current support, remove terms whose normalized-dictionary coefficients are below τ_{thr} in magnitude, and refit on the remaining support. This threshold–refit procedure is repeated until the support is unchanged or the prescribed maximum number of iterations is reached. After the final support is selected, the retained coefficients are recovered on the original candidate scale. In the reported experiments, we use $\lambda_{\mathrm{ridge}} = 10^{-10}$, $\tau_{\mathrm{thr}} = 2 \times 10^{-3}$, and at most five threshold–refit iterations. The reported e_{coef} is computed after back-scaling, over the complete six-entry coefficient vector.

Let $\mathbf{c}^\star$ be the true coefficient vector and $\widehat{\mathbf{c}}$ the identified vector. We report

$$e_{\mathrm{coef}} = \frac{\|\widehat{\mathbf{c}} - \mathbf{c}^\star\|_2}{\|\mathbf{c}^\star\|_2}, \qquad e_{\mathrm{roll}} = \frac{\|u_{\widehat{\mathbf{c}}}(\cdot, T_{\mathrm{roll}}) - u_{\mathrm{ref}}(\cdot, T_{\mathrm{roll}})\|_{L^2(\Omega)}}{\|u_{\mathrm{ref}}(\cdot, T_{\mathrm{roll}})\|_{L^2(\Omega)}}.$$

The first metric measures coefficient recovery over the complete candidate dictionary; the second measures whether the recovered law remains predictive far beyond the identification window.

The *ambient rank-matched* baseline replaces N_Ω by an ambient Fourier subspace of the same rank, while retaining the same sensors, state-fitting procedure and candidate mechanisms. It therefore controls for dimension but does not encode the target Dirichlet boundary. The *FD dense* baseline constructs pointwise spatial derivatives on the full geometry-matched masked grid and applies the same sparse-regression principle. It receives substantially denser spatial information than the proposed method. The *FD sparse* baseline uses only the same $N_{\mathrm{obs}} = 1000$ sensor values as Geometry-aware LegONet. The label *UNSTABLE* indicates divergence before $T_{\mathrm{roll}} = 1.0$.

Fig. 7a summarizes coefficient recovery and long-time prediction. On the peanut domain, the zero-noise experiment with $N_{\mathrm{obs}} = 1000$ gives

$$e_{\mathrm{coef}} = 6.89 \times 10^{-5}, \qquad e_{\mathrm{roll}} = 3.17 \times 10^{-5}.$$

At relative noise $\sigma = 10^{-2}$, the final rollout error remains 4.53×10^{-2}. Reducing the sensor count to $N_{\mathrm{obs}} = 240$ still gives $e_{\mathrm{roll}} = 1.83 \times 10^{-4}$.

For the channel, the zero-noise result is

$$e_{\mathrm{coef}} = 1.25 \times 10^{-2}, \qquad e_{\mathrm{roll}} = 3.44 \times 10^{-2}.$$

The recovered model remains predictive at low observation noise, whereas the coefficient and rollout errors increase at $\sigma = 10^{-2}$. The channel is more demanding because nonlinear transport must be identified from a short window on a thin curved geometry.

The representation ablation is decisive. The rank-matched ambient Fourier baseline gives coefficient errors 6.42×10^2 and 1.77×10^2 on the peanut and channel domains, respectively, with final rollout errors near one. The failure is not caused by rank alone: the ambient coordinates do not encode the target boundary.

Dense finite differences obtain competitive zero-noise rollout errors when given the full masked grid. Under the sparse $N_{\mathrm{obs}} = 1000$ sensor budget, however, the

peanut coefficient error rises to 9.33×10^2, and the channel rollout becomes unstable. The boundary-adapted block dictionary converts sparse state observations into stable mechanism features, avoiding the fragile derivative estimation required by finite-difference discovery on irregular domains.

Panels (b,c) of Fig. 7 show a representative zero-noise peanut rollout and its error history. The boundary-adapted identified model reproduces transport of the phase structure across the narrow neck, with a representative final error 1.50×10^{-4}. The rank-matched ambient Fourier and FD rollouts lose stability, with representative final errors 5.30×10^1 and 3.97×10^1, respectively. Panels (d,e) show the corresponding channel realization. The identified law transports the coherent structure through the oblique curved domain, with representative active-coefficient error 1.09×10^{-2} and final visual error 2.62×10^{-3}. The correct active support is recovered. The ambient Fourier and FD baselines fail early, with representative final errors 3.33×10^2 and 7.67, respectively. Note that these visualizations are single-seed examples and are not the means reported in panel (a).

Together, the coefficient, rollout and baseline diagnostics show that gLegONet turns sparse observations on unseen domains into stable mechanism features, enabling short-time identification to produce predictive long-time laws.

References

[1] Gronwall, T. H. Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Annals of Mathematics* **20**, 292–296 (1919).

[2] Pachpatte, B. G. *et al.* Inequalities for differential and integral equations. Tech. Rep., Academic press (1998).

[3] Zhang, J., Wang, Y. & Lin, G. LegONet: Plug-and-play structure-preserving neural operator blocks for compositional PDE learning. *arXiv preprint arXiv: 2603.07882* (2026).

[4] Rahman, M. A., Ross, Z. E. & Azizzadenesheli, K. U-NO: U-shaped neural operators. *arXiv preprint arXiv:2204.11127* (2022).