COORDINATE CLOSURE RECOVERS NONLINEAR COUPLING IN COARSE PDE PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Coarse PDE prediction can lose a nonlinear contribution even when current samples identify the structure that produces it. If a coherent coordinate has a continuation beyond the grid cutoff, projecting it before multiplication deletes cross interactions that return to represented modes. We introduce coordinate closure (C0): recover the coordinate from the current field, retain the visible residual, evaluate the finite returned-band interaction from analytic coordinate coefficients, and allocate this forcing between coordinate and residual dynamics. A Fourier identity localizes the required continuation and gives an exponential tail envelope. On Michelson–Sivashinsky fields, the complete transition beats exact-product controls on all 48 compatible outer fields. A cloned-state mask isolates a smaller 2.7–3.3% tail effect in the two clean low-band cells and its resolution and rank-zero exits. An independently implemented Benjamin–Ono transition reproduces tail sensitivity and the resolution exit under a different coordinate law. In an FNO with fixed parameter shapes, the interaction-aware feature and mask combination lowers pooled current-field forecast error by 14.9%. Seven of eight cellwise tests pass, but adding the interaction to history plus decomposition increases the descriptive pooled error, and resolved and coordinate-free fields favor the raw-input model. The learned result establishes conditional feature utility within this architecture; the deterministic mask isolates the tail mechanism. Together they support a conditional same-observation representation for coarse nonlinear prediction.

1 INTRODUCTION

Coarse nonlinear prediction can fail even when the current samples retain the relevant structured state. Consider a narrow coherent wave on a smooth background. Within a specified wave family, the samples identify the wave's phase and scale although its Fourier continuation extends beyond the grid cutoff. A band-limited transition that projects before multiplication removes cross terms between that continuation and the visible background. Some of those terms return to represented output modes. Both transitions receive the same samples, while one representation retains this nonlinear action (Figure 1).

This representation question matters for learned PDE models. Neural operators learn maps between sampled functions (Li et al., 2021; Lu et al., 2021; Kovachki et al., 2023), while learned discretizations and closure models estimate unresolved-scale effects (Bar-Sinai et al., 2019; Kochkov et al., 2021; Sanderse et al., 2025). Their accuracy depends on the available information and on how that information is presented to the nonlinear transition. Missing state may require history or finer sensing; operator aliasing may require a different discretization. Here the current field identifies a structured coordinate, but projection deletes its influence before prediction. The question is whether retaining that coordinate during nonlinear evaluation improves a model with the same input.

The setting is conditional on a recoverability prior: the accepted coordinate family specifies the continuation, and the encoder uses only the current field. The residual remains confined to observed non-Nyquist modes. The forecast is therefore a deterministic function of the original samples; a sufficiently expressive raw-grid network could represent the same map. We test whether exposing the interaction changes finite-data learning at fixed tensor shapes and parameter count. This separates representation from memory for partially observed dynamics (Chorin et al., 2002; Buitrago Ruiz et al., 2025) and from representation-consistent discretization (Bartolucci et al., 2023).

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

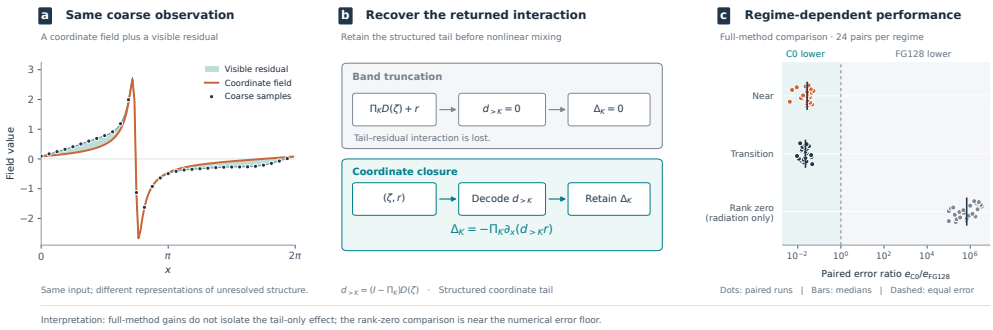

Figure 1: **Projection can delete an interaction recoverable from the same samples.** (a) C0 recovers a structured coordinate from all 128 samples of a frozen MS field. (b) Projection before multiplication removes the returned interaction Δ_K. (c) C0 has lower error than the exact-product FG128 control on all 48 compatible fields, whereas FG128 has lower error on all 24 coordinate-free fields. The cloned-state T0 mask, rather than the full C0–FG gap, isolates the tail-specific effect.

Write the represented field as $v = D(\boldsymbol{\zeta}) + r$, where $D(\boldsymbol{\zeta})$ is the coordinate decoder and r is a residual polynomial on the non-Nyquist band. Let $\boldsymbol{\Pi}_K$ project continuous fields onto $|k| \le K$. For a quadratic vector field $\mathbf{Q}(v, v)$, define the full cross interaction by $\mathbf{B}(d, r) = \mathbf{Q}(d, r) + \mathbf{Q}(r, d)$. Thus $\mathbf{B}(d, r) = -\partial_x(dr)$ for Michelson–Sivashinsky (MS) and $\mathbf{B}(d, r) = -2\partial_x(dr)$ for Benjamin–Ono (BO). With pure-coordinate dynamics supplied separately, truncation changes this cross interaction by

$$\Delta_K(\boldsymbol{\zeta}, r) := \boldsymbol{\Pi}_K \mathbf{B}(D(\boldsymbol{\zeta}), r) - \boldsymbol{\Pi}_K \mathbf{B}(\boldsymbol{\Pi}_K D(\boldsymbol{\zeta}), r) = \boldsymbol{\Pi}_K \mathbf{B}((\mathbf{I} - \boldsymbol{\Pi}_K) D(\boldsymbol{\zeta}), r)\,. \tag{1}$$

We call this term the *representation-closure defect*. Coordinate closure (C0) recovers the coordinate, retains the visible residual, computes the cross interaction from the analytic coordinate coefficients, and assigns the forcing between coordinate and residual dynamics. The coordinate's self-dynamics follow the known PDE-specific law. The construction continues only the coordinate coefficients required by the returned interaction and uses the current field throughout.

The evidence assigns distinct roles to analysis and controls. An exact identity localizes the returned interaction and yields a decay envelope. MS exact-product and time-refinement controls test the complete transition; a cloned-state mask isolates the smaller tail-specific contribution. A separate BO implementation tests the same mechanism under a different coordinate and evolution law. A learned comparison with fixed parameter shapes then tests feature and mask choices with decomposition and history. Resolution, coordinate-free inputs, and learned boundary fields provide explicit exits.

2 Related Work

Operator representation and discretization. FNO and DeepONet learn operators from discretized input–output pairs (Li et al., 2021; Lu et al., 2021). Nonlinear decoders represent solution sets on nonlinear manifolds (Lanthaler et al., 2023; Seidman et al., 2022), while coordinate-transform and particle operators alter the chart or basis used by the learned map (Gao et al., 2024; Liu et al., 2026; Li et al., 2026). Bartolucci et al. (2023) formalize consistency between continuous operators and their discrete realizations; CNO constructs representation-equivalent layers (Raonić et al., 2023). Coordinate closure addresses a related choice: which continuation of a recovered structure enters nonlinear evaluation. Our exact-product Galerkin control removes product aliasing for its represented polynomial, but retains sampling aliases in its raw-grid initialization. The coordinate model instead uses an explicit family to recover continuous coefficients before forming selected products.

Closure, nonlinear lifting, and memory. Learned discretizations, solver corrections, and multi-scale operators infer unresolved effects from coarse states (Bar-Sinai et al., 2019; Kochkov et al., 2021; Lütjens et al., 2022). Nonlinear Galerkin and approximate inertial-manifold methods reconstruct unresolved variables (Marion & Temam, 1989; Jolly et al., 1990); equation-free methods lift, evolve, and restrict (Kevrekidis et al., 2004). Coordinate closure is a structured lifting construction

with a specified coordinate–residual cross term and a PDE-specific latent update. Its contribution is the realization and controlled attribution of that term, rather than the general idea of reconstructing unresolved variables. Eliminating information can also introduce memory (Chorin et al., 2002). Buitrago Ruiz et al. (2025) demonstrate benefits from history in partially observed PDEs. Our instantaneous construction assumes recoverability from the current field, and its history comparison concerns one fixed architecture and distribution. These distinctions follow the separation of physical priors, discretization, and prediction objectives in closure modeling (Sanderse et al., 2025; Brunton & Kutz, 2024).

Structured recovery and coherent coordinates Matrix pencils, ESPRIT, and finite-rate-of-innovation methods identify compact exponential parameters under structured measurement assumptions (Sarkar & Pereira, 1995; Roy & Kailath, 1989; Vetterli et al., 2002), with stability depending on separation and noise (Batenkov & Yomdin, 2013). Moving-coordinate methods similarly use low-dimensional geometry to represent transported fields (Zhu et al., 2025; Taylor et al., 2025). Classical MS poles and BO travelling profiles provide the two families used here (Thual et al., 1985; Sivashinsky, 1977; Benjamin, 1967; Ono, 1975). Recovery identifies the coherent component; it does not determine whether its continuation matters to a forecast. The returned-band calculation and tail-removal experiments address that additional question.

3 Coordinate Closure: Recover, Retain, Interact

C0 is a deterministic transition on the current coarse field. At each step it recovers a compact coordinate, retains every visible non-Nyquist residual mode, evaluates the coordinate–residual interaction before truncation, and allocates the resulting forcing between the two state channels. It then advances both channels. Future observations and generator coordinates never enter the transition. Table 2 in Appendix A gives the complete input–output contract.

3.1 Coordinate and visible residual

The primary realization uses the periodic Michelson–Sivashinsky (MS) slope equation

$$\partial_t v + v\partial_x v = \nu\partial_{xx} v + \mathbf{H}\partial_x v, \qquad \widehat{\mathbf{H}\partial_x v}(k) = |k|\widehat{v}(k), \tag{2}$$

on $\mathbb{T}$. Let $\mathcal{G}_N = \{x_i = 2\pi i/N\}_{i=0}^{N-1}$ and $\mathbf{v}^n = (v(x_i, t_n))_{i=0}^{N-1} \in \mathbb{R}^N$. From only $(\mathbf{v}^n, \nu, N, \Delta t)$, the encoder returns

$$E_{N,\nu}(\mathbf{v}^n) = \big(\mathcal{Z}^n, \widehat{\mathbf{r}}^n\big), \qquad \mathcal{Z}^n = \{z_j^n = a_j^n + iy_j^n\}_{j=1}^{n_{\mathrm{p}}}. \tag{3}$$

All discrete coefficients use the forward normalization $[\mathbf{F}_N\mathbf{w}]_k = N^{-1}\sum_{j=0}^{N-1} w_j e^{-2\pi ijk/N}$, $k = 0, \ldots, N-1$. The coordinate decoder and visible residual are

$$p(x; \mathcal{Z}, \nu) = 2\Re\sum_{j=1}^{n_{\mathrm{p}}}(-\nu)\cot\left(\frac{x - z_j}{2}\right), \qquad \mathbf{r}^n = \mathbf{v}^n - \mathbf{S}_N p(\cdot; \mathcal{Z}^n, \nu). \tag{4}$$

At $N = 128$, a frozen alias-aware matrix pencil uses modes 13:40 to recover the roots $e^{-y_j - ia_j}$. C0 then retains every non-Nyquist Fourier coefficient of $\mathbf{r}^n$. Recovery gates and the grid-specific resolution windows are given in Appendices A.2 and B.2.1.

3.2 Closing and allocating the returned interaction

Set $K = N/2 - 1$. The reconstruction $r^n = \mathbf{I}_N\mathbf{r}^n$ retains only DFT modes $|k| \le K$ (Equation (12)). The MS transition evaluates the retained residual-channel nonlinear forcing

$$\widehat{\mathbf{q}}_k^n = -ik\left[\sum_{\ell=-K}^{K}\widehat{p}^n(k-\ell)\widehat{r}^n(\ell) + \frac{1}{2}\sum_{\ell=-K}^{K}\widehat{r}^n(\ell)\widehat{r}^n(k-\ell)\right], \qquad |k| \le K, \tag{5}$$

where out-of-band residual coefficients are zero, whereas the recovered pole law supplies $\widehat{p}^n(m)$ through $|m| \le 2K$. The pure-pole self-interaction and linear coherent evolution are carried by

the known pole law; Equation (5) contains the pole–residual and residual–residual terms left after subtracting that equation. This asymmetric continuation is the design decision forced by Equation (1): the coordinate is continued only inside products that return to represented outputs, while the residual remains on the observed band.

For MS, the omitted term has coefficients

$$\widehat{\Delta_K}(k) = -ik \sum_{\substack{|m|>K,\ |\ell|\le K \\ m+\ell=k}} \widehat{p}(m)\widehat{r}(\ell), \qquad |k| \le K. \tag{6}$$

If $|\widehat{p}(m)| \le c_p e^{-|m|y_{\min}}$ for $c_p \ge 0$, $y_{\min} > 0$, and the residual is supported on $|\ell| \le R \le K$, then

$$\|\Delta_K\|_2 \le K\sqrt{2R}\, c_p e^{-(K+1)y_{\min}} \sum_{|\ell|\le R} |\widehat{r}(\ell)|. \tag{7}$$

Only the coordinate shell $K < |m| \le K + R$ contributes, and $\widehat{\Delta_K}(k) = 0$ for $|k| \le K - R$. Thus a compact coordinate supplies a finite set of coefficients needed by the retained outputs without evolving an arbitrary fine-grid field. The bound gives an exponential envelope. A single-pole calculation in Appendix N realizes the scale $Ke^{-(K+1)y}$ and makes this distinction explicit.

The full-band Fourier–Galerkin control evolves $|k| \le K$ from the non-Nyquist reconstruction of the raw samples, with an exact projected product (Appendix A.2.2). This evaluates the observed-band product, while sampling aliases remain encoded in the observations. Its outputs lie in $\mathcal{V}_K$, the sampled real trigonometric polynomials with frequencies $|k| \le K$. C0 instead evaluates the recovered coordinate directly on the grid, so its output may extend outside $\mathcal{V}_K$. The full comparison therefore includes both transition and output-representation effects. The cloned-state T0 intervention isolates the returned-tail term.

The retained forcing must then be assigned to the two state channels. Let $\mathbf{J}_c^n \in \mathbb{R}^{N\times 2n_{\mathrm{p}}}$ contain the centered sampled phase and height derivatives of the decoded pole field, and let $\mathbf{q}_c^n$ be the centered forcing from Equation (5). If $n_{\mathrm{p}} = 0$, C0 skips the projection, takes $\mathbf{b}^n$ to be empty, and sets $\mathbf{q}_\perp^n = \mathbf{q}_c^n$. For $n_{\mathrm{p}} \ge 1$, it computes the deterministic ridge projection

$$\mathbf{b}^n = \arg\min_{\mathbf{b}\in\mathbb{R}^{2n_{\mathrm{p}}}} \frac{1}{N}\|\mathbf{q}_c^n - \mathbf{J}_c^n\mathbf{b}\|_2^2 + \rho_n\|\mathbf{b}\|_2^2, \qquad \mathbf{q}_\perp^n = \mathbf{q}_c^n - \mathbf{J}_c^n\mathbf{b}^n, \tag{8}$$

with the fixed scale $\rho_n = \rho\,\mathrm{tr}((\mathbf{J}_c^n)^\top\mathbf{J}_c^n/N)/(2n_{\mathrm{p}})$ and $\rho = 10^{-12}$. On the positive-rank branch, the tangent increment $\mathbf{b}^n$ augments the known MS pole velocity, while the ridge remainder $\mathbf{q}_\perp^n$ advances the residual with a midpoint exponential update. Every reported macro-step uses 16 substeps. The pole law, midpoint update, and mean gauge are specified in Appendix A.2.1. The optimization in Equation (8) is a deterministic partition of known forcing.

3.3 BO AND LEARNED REALIZATIONS

Periodic BO uses a different equation and coordinate family:

$$u_t = \partial_x\big(|\mathbf{D}|u - u^2\big), \qquad P_{a,y}(x) = \frac{1-e^{-2y}}{1-2e^{-y}\cos(x-a)+e^{-2y}}, \qquad \widehat{P}_{a,y}(k) = e^{-|k|y}e^{-ika}. \tag{9}$$

Here $\mathbf{D} = -i\partial_x$, and $|\mathbf{D}|$ is the Fourier multiplier with symbol $|k|$. Its raw-field encoder fits the sampled alias law for $\lambda = e^{-y-ia}$; its returned term is $\Delta_K^{\mathrm{BO}} = -2\mathbf{\Pi}_K\partial_x((\mathbf{I}-\mathbf{\Pi}_K)P_{a,y}r)$. The BO tangent directions and update are derived from this coordinate rather than copied from MS (Appendix A.3.1). Thus the two realizations share the recover–retain–close principle while using distinct coordinate formulas and evolution laws.

Coordinate features for a learned transition A frozen two-window coordinate encoder supplies the learned predictor with the sampled coordinate $\mathbf{p}^n$, residual $\mathbf{r}^n$, and returned-interaction field $\mathbf{d}^n = \mathbf{S}_N\Delta_K^n$. The matched experiment inserts normalized versions of these fields, the current raw field, and four past fields into a fixed eight-slot matrix $\mathbf{T}_a^n$. Arm a selects slots through a mask $\mathbf{m}_a^n$;

rejected coordinates disable the pole and interaction slots. A shared-token FNO architecture predicts a normalized increment and adds it to the current field,

$$\mathbf{v}_{\mathrm{pred}}^{n+1} = \mathbf{v}^n + s_v f_{\theta_a}(\mathbf{T}_a^n, \mathbf{m}_a^n, \widetilde{\nu}, \mathbf{c}_a^n), \tag{10}$$

where s_v is the fixed field scale, $\widetilde{\nu}$ is normalized viscosity, and $\mathbf{c}_a^n$ contains recovery confidences. All arms retain the same raw-field skip connection and have identical parameter shapes; each is trained separately. R activates one raw-field slot, S two coordinate and residual slots, and C adds a third interaction slot. H uses the current field and four past fields; HS and HC add the corresponding coordinate features to H. The comparison tests these feature and masking choices under fixed nominal parameter shapes; the active aggregation differs (Appendix E.1).

Evidence contract Four observations must agree for the representation claim to hold. The current raw field must identify an accepted coordinate without generator or future information. Its out-of-band continuation must produce a measurable returned interaction at the deployment cutoff. Removing that continuation, the visible residual, or tangent assignment must change the forecast in the predicted direction. Same-grid time refinement and exact observed-band convolution test numerical explanations for the complete-transition gain, while rank-zero and resolved-tail fields test the predicted exit of the tail-specific effect.

This contract distinguishes a returned-band defect from high-frequency energy alone. Equation (6) requires a visible residual mode to return a coordinate coefficient to $|k| \le K$; Equation (7) predicts contraction as $(K+1)y_{\min}$ grows; and rank-zero recovery sets the coordinate factor to zero. Numerical controls test the complete transition, but only the cloned-state mask attributes its tail component.

4 Experiments and Results

The experiments ask three questions in order. First, does the interaction removed by coarse projection exist, and can it be recovered from the current field? Second, does retaining that interaction causally improve a forecast, rather than merely change the time integrator, initialization, or output space? Third, does the same calculation help a parameter-matched learned operator? The first two questions concern the deterministic transition; the third concerns representation within one fixed neural architecture.

Common protocol All deployment fields are real and 2π periodic and use $N = 128$ endpoint-free observations unless a resolution ladder is stated explicitly. The MS outer panels use a two-thirds-dealiased 1,024-point ETDRK4 reference checked at 2,048 points; the cloned-state tail panel uses a 2,048-point reference checked at 4,096 points. BO uses an independently implemented 1,024-point, twofold-zero-padded ETDRK4 reference. For field b, the primary deterministic endpoint is

$$e_b = \frac{1}{5}\sum_{h=1}^{5} \frac{\|\mathbf{v}_{b,h}^{\mathrm{pred}} - \mathbf{v}_{b,h}\|_2}{\|\mathbf{v}_{b,h}\|_2}. \tag{11}$$

The matched learned study averages its first ten of 20 saved horizons. In all uncertainty calculations, the independent unit is a physical field or trajectory, not a horizon, observation view, or training seed. Full solvers, registries, and bootstrap procedures are given in Appendix B. PS64 refines time stepping, $K16$ restricts residual bandwidth, NT removes tangent allocation, and FG64/FG128 use exact projected convolution on all observed non-Nyquist modes. T0 masks only the coordinate tail in that product.

4.1 Recovery of the returned interaction

The B0–D1 diagnostic measures the omitted interaction before rollout error is introduced. It follows 18 fixed MS fields across $N \in \{64, 128, 256, 512\}$. The relative omitted feedback decreases as the resolved-tail index $(K+1)y_{\min}$ grows: 17 of 18 fieldwise ladders are non-increasing, and the mean within-field log slope is -0.805 with 95% field-bootstrap interval $[-0.859, -0.747]$. This is the resolution signature predicted by Equation (7); the fitted slope itself is descriptive.

Existence alone is insufficient if the interaction cannot be obtained from the deployed observation. On the 24 declared under-resolved near and transition records, coupling computed from the recovered

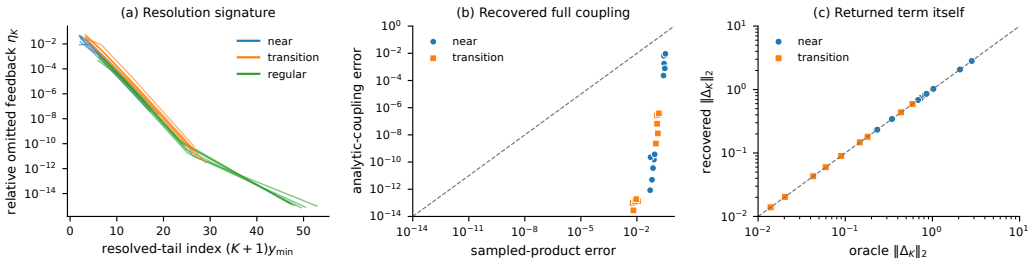

Figure 2: **The returned interaction is measurable from the current coarse field.** (a) Each path follows one of 18 fixed physical fields across four grids; the omitted feedback contracts with the resolved-tail index. (b) On the declared under-resolved records, recovered analytic coupling is more accurate than sampled-grid multiplication; the dashed line marks equal error. (c) Reconstructed returned-term norms track the generator oracle; the median and worst relative field errors are reported in the text. Colors denote field geometry and the dashed line in (c) marks equality.

analytic coordinate is more accurate than native-grid multiplication in every case. We further compare the returned term itself with the generator-level oracle Δ_K. On the material $N \in \{64, 128\}$ near and transition records, the median relative field error is 7.3×10^{-10} and the maximum is 1.96%. The norms track the identity line, while the maximum records imperfect recovery in the hardest case. These direct-tail diagnostics were reconstructed on the existing registry; their provenance and complete records are given in Appendix O.

4.2 Forecast effects and causal attribution

Initialization and sampled-output comparisons. The four-variant diagnostic uses 24 fields in each stratum, with common observations, references, and five horizons. FG-raw initializes exact-product Galerkin evolution from the raw sample DFT; FG-rec uses recovered continuous coordinate and residual coefficients. C0-proj projects each decoded C0 forecast onto the same non-Nyquist sample subspace as FG; C0-full retains the complete decoded sample vector. This diagnostic was reconstructed from the supplied fixed runner and existing field registries after its original result file was found missing (Appendix O).

Recovered initialization gives geometric-mean FG-rec/FG-raw error ratios 1.012 and 1.004 in the near and transition strata. C0-proj is lower than FG-rec on all 48 compatible fields, with ratios 0.294 and 0.291. These results establish a predictive advantage within a common *sample* subspace. Projection after sampling retains aliases of the coordinate's continuous high modes; it does not isolate continuous low-mode dynamics. C0-full/C0-proj ratios of 0.079 and 0.073 measure the benefit of restoring the single discrete Nyquist direction at fixed forecasts, not the whole decoder contribution (Appendix O.1). T0 supplies the separate tail-specific intervention. All 24 radiation-only fields reverse the C0/FG ordering, with FG at the numerical floor.

Cloned-state intervention on the unresolved tail. The complete-transition comparison above does not by itself identify the specific term in Equation (6). T0 supplies that intervention: it starts from the same recovered state as C0 and keeps the visible residual, tangent allocation, coordinate update rule, decoder, and integration schedule fixed, but masks coordinate coefficients $K < |m| \le 2K$ inside the cross product. Removing those coefficients increases clean low-band error by 2.7–3.3% relative to C0 in the near and transition cells. The contrast contracts to numerical equality at $N \ge 256$ and is identically zero when recovery returns rank zero. Overlap-band and 0.5%-noise cells lie near equality even though every declared observation view passes recovery. Hence coordinate recovery and returned-interaction materiality are separate conditions.

The remaining controls address numerical alternatives. PS16 and PS64 share a same-grid time-refinement plateau, while $K16$ and NT show that visible residual modes and tangent allocation contribute to the complete transition. FG64 and FG128 are numerically indistinguishable on compatible fields, while C0 wins all 48 comparisons against FG128. An independently implemented BO study changes the PDE, coordinate family, decoder, tangent basis, and update law. Its raw-field C0–SP

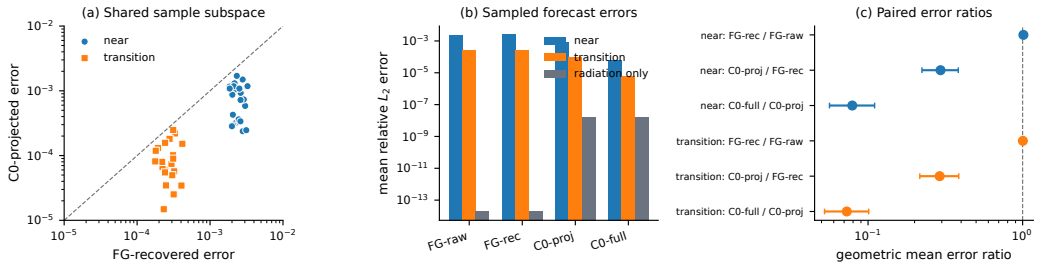

Figure 3: **Initialization and sample-space projection control different parts of the forecast comparison.** (a) C0-proj and FG-rec share the non-Nyquist sample subspace; every compatible field favors C0-proj. (b) Mean sampled errors for the four variants; coordinate-free fields reverse the ordering. (c) Geometric-mean paired ratios and 95% field-bootstrap intervals, conditional on the reconstructed diagnostic. High continuous modes can alias into C0-proj, so (a) is not a continuous low-mode dynamics comparison.

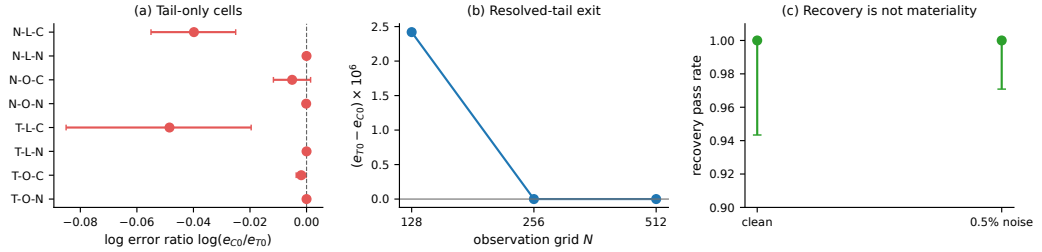

Figure 4: **A cloned-state mask isolates the smaller unresolved-tail effect.** (a) Paired $\log(e_{\mathrm{C0}}/e_{\mathrm{T0}})$ estimates across geometry, residual spectrum, and observation view; negative values favor C0. N/T denote near/transition geometry, L/O low/overlap residual spectra, and C/N clean/noisy observations. (b) The same-state contrast disappears as the tail is resolved. (c) Clean and noisy observations both pass recovery, showing that successful recovery alone does not imply a detectable endpoint effect. Bars show the recorded 95% intervals.

contrast favors C0 in the active near and transition strata and returns to equality after the coordinate tail is resolved. These controls and the complete BO display are reported in Appendices C.2 and D.1, Figure 9, and Table 8. The BO mean errors are 1.388×10^{-3} for C0 versus 1.602×10^{-3} for SP near the axis, and 7.71×10^{-4} versus 8.27×10^{-4} in transition. Its resolution ladder is an offline generator-coordinate diagnostic, not a recovered-coupling accuracy test.

4.3 Matched learned-operator comparison

The learned experiment fixes one eight-slot FNO with 100,657 real parameters. R receives the current raw field, S receives the recovered coordinate and visible residual, and C adds the returned-interaction token to S. H receives the current field and four past fields; HS and HC add the corresponding coordinate features to H. Coordinate tokens are deterministic functions of the current observation, so R, S, and C have the same information set. All arms have identical learned tensor shapes and are trained separately with three confirmation seeds.

The outer registry contains 512 physical trajectories in four geometry–spectrum cells, with clean and noisy views forming eight equally weighted analysis cells. Pooled superiority uses paired one-sided 99% upper log bounds below zero; the all-cell gate requires eight Holm-adjusted tests to pass. C reduces pooled error by 14.9% relative to R (C/R $= 0.851$), and C/S $= 0.665$ shows that decomposition alone does not explain the gain. The positive comparison concerns the feature and mask combination: R/S/C activate one/two/three slots, and R's single-slot softmax makes its query/key pathways inactive (Appendix E.1). Equal nominal parameter counts therefore do not isolate the physical tail mechanism. H does not improve over R under its prespecified superiority test, and HC does not improve over H; failure of these tests is not interpreted as equivalence. Descriptively,

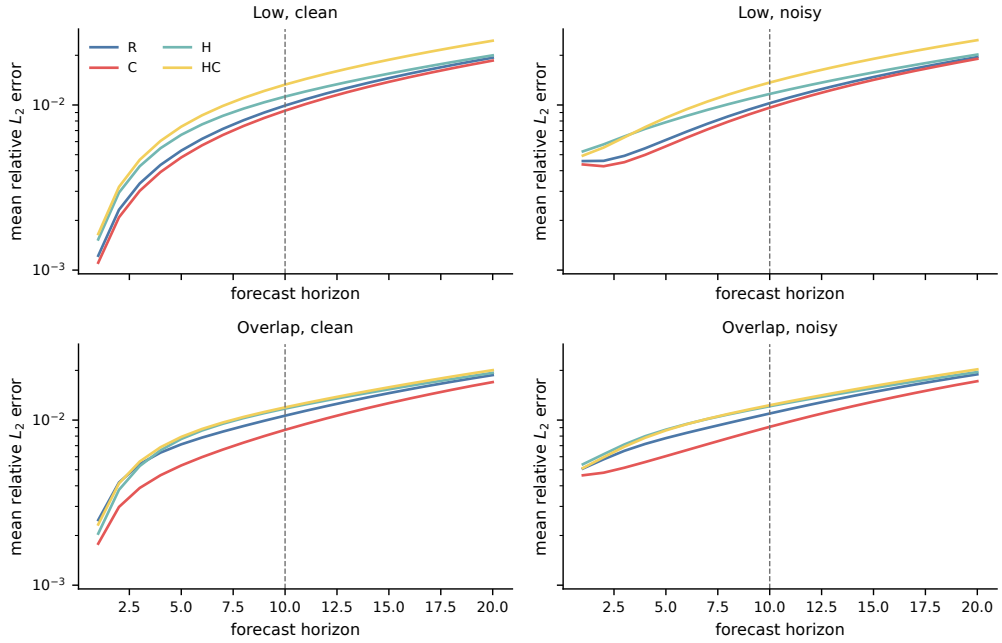

Figure 5: **The interaction-aware feature and mask combination improves the pooled current-field forecast.** Curves average three confirmation runs within each trajectory and then average 256 physical trajectories in each spectrum–view panel. R uses the current raw field, C adds the current coordinate decomposition and interaction, H uses the current field and four past raw fields, and HC combines history with the coordinate features. All arms share parameter shapes. The vertical rule ends the ten-horizon primary window; horizons 11–20 are descriptive.

HS/H $= 0.808$ but HC/HS $= 1.311$: adding the interaction to history plus decomposition reverses the favorable current-field ordering.

Table 1: **Primary matched learned contrasts.** Ratios compare pooled mean errors; values below one favor the numerator. The one-sided 99% upper log bound must be below zero for superiority. C/H instead uses the frozen noninferiority margin $\log(1.05)$. Models share parameter shapes, not weights.

Contrast	Error ratio	99% upper log bound	Decision
C/R	0.8510	−0.1475	pass
C/S	0.6648	−0.3926	pass
H/R	1.1433	0.1463	fail
HC/H	1.0592	0.0756	fail
C/H	0.7444	−0.2807	noninferiority pass

Prior-asymmetric replacement. C0 is lower on every C1 outer trajectory and every C2 trajectory at the tested 64–1,024-trajectory FNO budgets, while FNO improves with more data (Figures 12 and 13). These observation-matched comparisons use different physical priors and costs; they support finite-data replacement claims, not a parameter-efficiency or asymptotic advantage.

Scope and failure modes. Seven of eight C/R cells pass; the transition/low-band/clean cell remains near equality, so the all-cell gate fails. Resolved and coordinate-free inventories favor R (C/R $= 2.4$–17.7; no paired C wins). Zeroing the interaction does not recover R's trained map; no safe fallback or selector is validated. C takes about twice R's training time. Evidence covers compatible one-dimensional periodic families and short forecasts. Full cellwise results, boundary tests, costs, and failed KdV/Allen–Cahn transfer gates appear in Appendices A.3, E and K.

5 CONCLUSION

Coarse samples can identify structure whose nonlinear influence is nevertheless lost by projection. C0 retains that structure as a coordinate and evaluates its finite returned-band interaction with the visible residual. Exact identities, MS numerical controls, a cloned-state tail intervention, and an independent BO realization separate the complete transition from its smaller tail component. An FNO with fixed parameter shapes benefits from the interaction-aware feature and mask combination at fixed observations. This learned gain depends on the input contract: the history-plus-decomposition arm and distribution-boundary fields reverse the favorable ordering. Sample-space projection controls the Nyquist output direction while retaining aliases of continuous coordinate tails; T0 supplies the tail-specific attribution. The evidence supports recoverable coordinates with material returned interactions in the declared families, with learned utility conditional on architecture and training.

REPRODUCIBILITY STATEMENT

The supplementary package provides paper-asset generators, frozen configurations and results, source and result hashes, a renderer, and a paper-asset notebook. The missing direct-tail and four-variant records were reconstructed on existing registries and are labeled separately (Appendix O). Appendix F maps claims to code and records; Appendix B documents splits, independent units, uncertainty, stopping rules, and learned selection and retraining seeds. The matched coordinate–history record is hash-pinned, and its primary table is generated directly from that record. The notebook rebuilds the displayed evidence from frozen records while preserving the recorded training studies. The figure-generation and final-size QA procedure is versioned with the evidence.

AI USE STATEMENT

Generative AI tools assisted with research brainstorming; methodology and experiment design; code development and review; mathematical checking; literature search; figure preparation; interpretation of saved results; and manuscript drafting and editing. No generative AI model is part of the proposed method or evaluated system. The authors remain responsible for the paper's claims, code, and artifacts.

REFERENCES

Yohai Bar-Sinai, Stephan Hoyer, Jason Hickey, and Michael P. Brenner. Learning data-driven discretizations for partial differential equations. *Proceedings of the National Academy of Sciences*, 116(31):15344–15349, 2019. doi: 10.1073/pnas.1814058116.

Francesca Bartolucci, Emmanuel de Bézenac, Bogdan Raonić, Roberto Molinaro, Siddhartha Mishra, and Rima Alaifari. Representation equivalent neural operators: A framework for alias-free operator learning. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-3051. URL https://papers.neurips.cc/paper_files/paper/2023/hash/dc35c593e61f6df62db541b976d09dcf-Abstract-Conference.html.

Dmitry Batenkov and Yosef Yomdin. On the accuracy of solving confluent Prony systems. *SIAM Journal on Applied Mathematics*, 73(1):134–154, 2013. doi: 10.1137/110836584.

T. Brooke Benjamin. Internal waves of permanent form in fluids of great depth. *Journal of Fluid Mechanics*, 29(3):559–592, 1967. doi: 10.1017/S002211206700103X.

Steven L. Brunton and J. Nathan Kutz. Promising directions of machine learning for partial differential equations. *Nature Computational Science*, 4:483–494, 2024. doi: 10.1038/s43588-024-00643-2.

Ricardo Buitrago Ruiz, Tanya Marwah, Albert Gu, and Andrej Risteski. On the benefits of memory for modeling time-dependent PDEs. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/89379d5fc6eb34ff98488202fb52b9d0-Abstract-Conference.html.

Alexandre J. Chorin, Ole H. Hald, and Raz Kupferman. Optimal prediction with memory. *Physica D: Nonlinear Phenomena*, 166(3–4):239–257, 2002. doi: 10.1016/S0167-2789(02)00446-3.

Wenhan Gao, Ruichen Xu, Hong Wang, and Yi Liu. Coordinate transform Fourier neural operators for symmetries in physical modelings. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=pMD7A77k3i.

Michael S. Jolly, Ioannis G. Kevrekidis, and Edriss S. Titi. Approximate inertial manifolds for the Kuramoto–Sivashinsky equation: Analysis and computations. *Physica D: Nonlinear Phenomena*, 44(1–2):38–60, 1990. doi: 10.1016/0167-2789(90)90046-R.

Ioannis G. Kevrekidis, C. William Gear, and Gerhard Hummer. Equation-free: The computer-aided analysis of complex multiscale systems. *AIChE Journal*, 50(7):1346–1355, 2004. doi: 10.1002/aic.10106.

Dmitrii Kochkov, Jamie A. Smith, Ayya Alieva, Qing Wang, Michael P. Brenner, and Stephan Hoyer. Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences*, 118(21):e2101784118, 2021. doi: 10.1073/pnas.2101784118.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL https://jmlr.org/papers/v24/21-1524.html.

Samuel Lanthaler, Roberto Molinaro, Patrik Hadorn, and Siddhartha Mishra. Nonlinear reconstruction for operator learning of PDEs with discontinuities. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=CrfhZAsJDsZ.

Zhihao Li, Yu Feng, Zhilu Lai, and Wei Wang. From basis to basis: Gaussian particle representation for interpretable PDE operators, 2026. URL https://arxiv.org/abs/2602.21551.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Chaoyu Liu, Zhonghao Li, Gaohang Chen, Zakhar Shumaylov, Zhongying Deng, Qian Zhang, Zhonghua Qiao, and Carola-Bibiane Schönlieb. Adaptive coordinate transforms for neural operators, 2026. URL https://arxiv.org/abs/2605.06203.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3:218–229, 2021. doi: 10.1038/s42256-021-00302-5.

Björn Lütjens, Catherine H. Crawford, Campbell D. Watson, Christopher Hill, and Dava Newman. Multiscale neural operator: Learning fast and grid-independent PDE solvers, 2022. URL https://arxiv.org/abs/2207.11417.

Martine Marion and Roger Temam. Nonlinear Galerkin methods. *SIAM Journal on Numerical Analysis*, 26(5):1139–1157, 1989. doi: 10.1137/0726063.

Hiroaki Ono. Algebraic solitary waves in stratified fluids. *Journal of the Physical Society of Japan*, 39(4):1082–1091, 1975. doi: 10.1143/JPSJ.39.1082.

Bogdan Raonić, Roberto Molinaro, Tim De Ryck, Tobias Rohner, Francesca Bartolucci, Rima Alaifari, Siddhartha Mishra, and Emmanuel de Bézenac. Convolutional neural operators for robust and accurate learning of PDEs. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/f3c1951b34f7f55ffaecada7fde6bd5a-Abstract-Conference.html.

Richard Roy and Thomas Kailath. ESPRIT—estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(7):984–995, 1989. doi: 10.1109/29.32276.

Benjamin Sanderse, Panos Stinis, Romit Maulik, and Shady E. Ahmed. Scientific machine learning for closure models in multiscale problems: A review. *Foundations of Data Science*, 7(1):298–337, 2025. doi: 10.3934/fods.2024043.

Tapan K. Sarkar and O. Pereira. Using the matrix pencil method to estimate the parameters of a sum of complex exponentials. *IEEE Antennas and Propagation Magazine*, 37(1):48–55, 1995. doi: 10.1109/74.370583.

Jacob H. Seidman, Georgios Kissas, Paris Perdikaris, and George J. Pappas. NOMAD: Nonlinear manifold decoders for operator learning. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://openreview.net/forum?id=5OWV-sZvMl.

G. I. Sivashinsky. Nonlinear analysis of hydrodynamic instability in laminar flames—I. derivation of basic equations. *Acta Astronautica*, 4(11–12):1177–1206, 1977. doi: 10.1016/0094-5765(77)90096-0.

Seth Taylor, Alex Bihlo, and Jean-Christophe Nave. Diffeomorphic neural operator learning, 2025. URL https://arxiv.org/abs/2508.06690.

Olivier Thual, Uriel Frisch, and Michel Hénon. Application of pole decomposition to an equation governing the dynamics of wrinkled flame fronts. *Journal de Physique*, 46(9):1485–1494, 1985. doi: 10.1051/jphys:019850046090148500.

Martin Vetterli, Pina Marziliano, and Thierry Blu. Sampling signals with finite rate of innovation. *IEEE Transactions on Signal Processing*, 50(6):1417–1428, 2002. doi: 10.1109/TSP.2002.1003065.

Yameng Zhu, Jingrun Chen, and Weibing Deng. R-adaptive DeepONet: Learning solution operators for PDEs with discontinuous solutions using an R-adaptive strategy. *Numerical Mathematics: Theory, Methods and Applications*, 18(1):31–65, 2025. doi: 10.4208/nmtma.OA-2024-0124.

Table 2: **One C0 transition.** Each operation uses the current field and fixed PDE parameters.

Operation	Input	Output
Recover	raw modes 13:40	accepted $\mathcal{Z}^n$ or rank zero
Retain	$\mathbf{v}^n - \mathbf{S}_N p(\mathcal{Z}^n)$	all represented non-Nyquist modes
Close	coordinate–residual coefficients	represented forcing
Allocate	forcing and decoder Jacobian	tangent and residual increments
Advance	both increments	$\mathbf{v}^{n+1}$

A Supplementary method details

This appendix expands the transition and the evidence contracts used in the main paper. It is organized so that a reader can distinguish a mathematical identity, a deployment computation, and an offline audit quantity.

A.1 Returned-band closure identity

Let $\mathbf{S}_N$ denote sampling on the deployment grid, $\mathbf{\Pi}_K$ the represented continuous Fourier projection, and $\mathbf{I}_N$ the non-Nyquist reconstruction operator. For even N, the reconstruction operator is

$$(\mathbf{I}_N \mathbf{w})(x) = \sum_{k=-K}^{K} [\mathbf{F}_N \mathbf{w}]_{k \bmod N} e^{ikx}, \qquad K = N/2 - 1. \tag{12}$$

It discards the discrete Nyquist coefficient. Thus $\mathbf{S}_N \mathbf{I}_N$ is an orthogonal sample-space projection, not the identity on arbitrary grid vectors. Let $d = D(\boldsymbol{\zeta})$ be a decoded coordinate field and $\mathbf{r}_N$ the grid residual left after recovery. For a bilinear PDE operator $\mathbf{B}$, bilinearity gives

$$\begin{aligned} \mathbf{S}_N \mathbf{\Pi}_K \mathbf{B}\big(d, \mathbf{I}_N \mathbf{r}_N\big) = \mathbf{S}_N \mathbf{\Pi}_K \mathbf{B}\big(\mathbf{\Pi}_K d, \mathbf{I}_N \mathbf{r}_N\big) \\ + \mathbf{S}_N \mathbf{\Pi}_K \mathbf{B}\big((\mathbf{I} - \mathbf{\Pi}_K) d, \mathbf{I}_N \mathbf{r}_N\big). \end{aligned} \tag{13}$$

The second line is the returned-band term. It can be nonzero even though its coordinate factor lies outside the output band: a high coordinate coefficient and a visible residual coefficient may sum to a retained Fourier index. For MS, $\mathbf{B}(p, r) = -\partial_x(pr)$, so the second line is precisely Equation (6). For BO it carries the factor two in the quadratic cross term. This identity specifies what C0 evaluates; it does not assert that every coarse nonlinear operator has a useful coordinate continuation. An unpadded native-grid pointwise product is an alias-affected surrogate for this projected continuous product; the zero-padded Galerkin control is not.

For residual support $|\ell| \le R$, the MS term at $|k| \le K$ is

$$\widehat{d}_K(k) = -ik \sum_{|\ell| \le R} \chi_{\{|k-\ell|>K\}} \widehat{p}(k-\ell)\widehat{r}(\ell). \tag{14}$$

Here $\chi_{\{\cdot\}}$ denotes the scalar indicator function. It vanishes for $|k| \le K - R$. The exponential pole spectrum bounds every remaining factor by $c_p e^{-(K+1)y_{\min}}$, yielding Equation (7). The relation explains the experimental choice to vary pole height, residual band, and grid resolution separately.

A.2 Raw-grid recovery and tangent assignment

MS recovery uses a fixed alias-aware matrix pencil on the raw field. The encoder estimates the roots $e^{-y_j - ia_j}$ from a Hankel system built from high Fourier modes, applies the declared mirror-alias correction, and accepts the prescribed rank only when the numerical gates pass. It then decodes $p(\mathcal{Z}, \nu)$ and defines $\mathbf{r} = \mathbf{v} - \mathbf{S}_N p(\cdot; \mathcal{Z}, \nu)$. The recovery audit never supplies a generator coordinate to this procedure.

After analytic coupling, a finite-grid Jacobian $\mathbf{J}$ contains the decoded MS coordinate directions. Define $\mathbf{C}_N = \mathbf{I} - N^{-1}\mathbf{1}\mathbf{1}^\top$, $\mathbf{J}_c = \mathbf{C}_N\mathbf{J}$, and $\mathbf{f}_c = \mathbf{C}_N\mathbf{f}$. For $n_{\mathrm{p}} \geq 1$, C0 uses

$$\begin{aligned} \alpha &= \frac{\mathrm{tr}(\mathbf{J}_c^\top \mathbf{J}_c/N)}{2n_{\mathrm{p}}}, & \rho &= 10^{-12}, \\ \delta\boldsymbol{\zeta} &= \left(\frac{\mathbf{J}_c^\top \mathbf{J}_c}{N} + \rho\alpha\mathbf{I}\right)^{-1} \frac{\mathbf{J}_c^\top \mathbf{f}_c}{N}, & & \\ \mathbf{f}_{\mathrm{tan}} &= \mathbf{J}_c\,\delta\boldsymbol{\zeta}, & \mathbf{f}_{\mathrm{rem}} &= \mathbf{f}_c - \mathbf{f}_{\mathrm{tan}}. \end{aligned} \tag{15}$$

For $n_{\mathrm{p}} = 0$, the solve is skipped and $\mathbf{f}_{\mathrm{rem}} = \mathbf{f}_c$. Here $\mathbf{f}$ is the represented analytic forcing. The tangent component updates the coordinate and the ridge remainder updates radiation; exact orthogonality is recovered when $\rho = 0$. NT is informative because it holds the recovered coordinate, residual bandwidth, and analytic product fixed while changing only this assignment. The fitted Jacobian is not used to infer an unobserved target; it partitions a forcing already determined by the current raw field and the declared PDE.

A.2.1 MS DEPLOYMENT UPDATE

The conjugate-completed pole law is

$$\dot{z}_j^{\mathrm{MS}} = -\nu \sum_{\substack{z_\ell \in \mathcal{Z} \cup \overline{\mathcal{Z}} \\ z_\ell \neq z_j}} \cot\left(\frac{z_j - z_\ell}{2}\right) - i. \tag{16}$$

Let $\mathbf{f}_\xi$ interleave its real and imaginary parts, let $\mathbf{L}_\nu = \mathrm{diag}_k(|k| - \nu k^2)$, and let $\boldsymbol{\Phi}_1$ apply $\varphi_1(z) = (e^z - 1)/z$ diagonally, with $\varphi_1(0) = 1$. One midpoint–exponential substep is

$$\begin{aligned} \boldsymbol{\xi}^{s+\frac{1}{2}} &= \boldsymbol{\xi}^s + \frac{h}{2}(\mathbf{f}_\xi(\mathcal{Z}^s, \nu) + \mathbf{b}^s), \\ \widehat{\mathbf{r}}^{s+\frac{1}{2}} &= e^{\frac{h}{2}\mathbf{L}_\nu}\widehat{\mathbf{r}}^s + \frac{h}{2}\boldsymbol{\Phi}_1\left(\frac{h}{2}\mathbf{L}_\nu\right)\widehat{\mathbf{q}}_\perp^s, \\ \boldsymbol{\xi}^{s+1} &= \boldsymbol{\xi}^s + h(\mathbf{f}_\xi(\mathcal{Z}^{s+\frac{1}{2}}, \nu) + \mathbf{b}^{s+\frac{1}{2}}), \\ \widehat{\mathbf{r}}^{s+1} &= e^{h\mathbf{L}_\nu}\widehat{\mathbf{r}}^s + h\boldsymbol{\Phi}_1(h\mathbf{L}_\nu)\widehat{\mathbf{q}}_\perp^{s+\frac{1}{2}}. \end{aligned} \tag{17}$$

At encoding, the implementation stores $m_r = \Re[\widehat{\mathbf{r}}]_0$ and restores this same residual coefficient after every midpoint and full stage. It does not constrain the sampled mean of $p + r$. Each reported macro-step uses $h = \Delta t/16$ with $\Delta t = 0.01$.

A.2.2 EXACT OBSERVED-BAND CONTROL

With $K = N/2 - 1$, the control initializes its coefficients from the discrete Fourier transform of the raw samples, discards the Nyquist coefficient, and sets out-of-band coefficients to zero. It therefore starts from $\mathbf{I}_N\mathbf{S}_N v^0$, not $\boldsymbol{\Pi}_K v^0$. The Fourier–Galerkin control solves

$$\frac{\mathrm{d}}{\mathrm{d}t}\widehat{v}_k = (|k| - \nu k^2)\widehat{v}_k - \frac{ik}{2}\sum_{\ell=-K}^{K} \widehat{v}_\ell \widehat{v}_{k-\ell}, \qquad |k| \leq K. \tag{18}$$

Twofold zero padding evaluates the projected quadratic sum without product aliasing; it does not remove aliases in the sampled initial state. The output is this polynomial evaluated on the observation grid, hence an element of $\mathcal{V}_K = \mathbf{S}_N\mathcal{T}_K$, where $\mathcal{T}_K$ is the space of real trigonometric polynomials with frequencies $|k| \leq K$. Because the discrete Fourier transform is orthogonal on the N-point grid and $\mathcal{V}_K$ is spanned by its modes $|k| \leq K$, the attainability floor $e_\perp = \mathrm{dist}(\mathbf{S}_N v, \mathcal{V}_K)/\|\mathbf{S}_N v\|_2$ used in Appendix C.3 is the Nyquist component of the sampled dense truth. It is a lower bound for every output in this sample subspace when evaluated against the same reference. The saved floor runner regenerates reference solutions but no candidate forecasts; Appendix P states the numerical alignment caveat.

A.3 ALIASING AND EXECUTION COST

A.3.1 BO DEPLOYMENT

With $\lambda = e^{-y-ia}$, the BO encoder uses the sampled law

$$\widetilde{P}_k^{(N)} = \frac{\lambda^k}{1-\lambda^N} + \frac{\overline{\lambda}^{N-k}}{1-\overline{\lambda}^N}, \qquad 1 \le k < N/2. \tag{19}$$

It sums the positive and negative geometric tails that alias into the same discrete coefficient. This distinction matters near the grid cutoff: fitting a continuous coefficient formula directly to a sampled DFT changes the inverse problem. The BO outer transition nevertheless receives only the raw grid and its frozen normalization.

The same alias sum fixes the sampled zero coefficient and background:

$$\widetilde{P}_0^{(N)} = 1 + \frac{\lambda^N}{1-\lambda^N} + \frac{\overline{\lambda}^N}{1-\overline{\lambda}^N}, \qquad \beta = \Re\Big([\mathbf{F}_N \mathbf{u}]_0 - \widetilde{P}_0^{(N)}\Big). \tag{20}$$

The residual is the normalized DFT of $\mathbf{u} - \beta\mathbf{1} - \mathbf{S}_N P_{a,y}$, masked to the declared non-Nyquist band, with its zero and Nyquist coefficients set to zero.

The Poisson identity $P_{a,y}^2 = |\mathbf{D}|P_{a,y} + \coth(y)P_{a,y}$ gives $\partial_x(|\mathbf{D}|(\beta + P_{a,y}) - (\beta + P_{a,y})^2) = -(\coth(y) + 2\beta)\partial_x P_{a,y}$. For $K = N/2 - 1$ and $|k| \le K$, the retained residual forcing used by the implementation is

$$\begin{aligned}\widehat{f}_{\mathrm{BO}}(k) = ik\Bigg[&|k|\widehat{r}(k) - 2\beta\widehat{r}(k) \\ &- 2\sum_{\ell=-K}^{K} \widehat{P}_{a,y}(k-\ell)\widehat{r}(\ell) - \sum_{\ell=-K}^{K} \widehat{r}(\ell)\widehat{r}(k-\ell)\Bigg],\end{aligned} \tag{21}$$

where out-of-band residual coefficients vanish and the analytic Poisson coefficients are evaluated through $|k| = 2K$. Writing $(\delta a, \delta y)$ for the real coefficient-space least-squares increment, the implementation solves

$$(\delta a, \delta y) = \arg\min_{(b_1,b_2)\in\mathbb{R}^2} \sum_{|k|\le K} \left|\widehat{f}_{\mathrm{BO}}(k) - b_1\partial_a\widehat{P}_{a,y}(k) - b_2\partial_y\widehat{P}_{a,y}(k)\right|^2, \tag{22}$$

with $\partial_a\widehat{P}_k = -ik\widehat{P}_k$ and $\partial_y\widehat{P}_k = -|k|\widehat{P}_k$. The real Gram matrix must have finite condition number at most 10^8. Set $\widehat{\mathbf{p}}_{a,y} := (\widehat{P}_{a,y}(k))_{|k|\le K}$. The BO split is

$$\dot{a} = \coth(y) + 2\beta + \delta a, \qquad \dot{y} = \delta y, \qquad \dot{\beta} = 0, \qquad \partial_t\widehat{\mathbf{r}} = \widehat{\mathbf{f}}_{\mathrm{BO}} - \delta a\,\partial_a\widehat{\mathbf{p}}_{a,y} - \delta y\,\partial_y\widehat{\mathbf{p}}_{a,y}. \tag{23}$$

BO C0 and its SP, $K16$, and NT controls use 24 classical RK4 substeps per 0.003 macro-step. The fixed-grid PS16 and PS64 controls use 16 and 64 substeps, respectively.

A.3.2 EXECUTION COST

For MS, a pencil sequence of length L forms an $L/2$-by-$L/2$ Hankel matrix and uses a full SVD. One C0 encoding therefore costs $O(N\log N + L^3 + Nn_{\mathrm{p}})$; $L = 28$ and $n_{\mathrm{p}} \le 2$ at the reported $N = 128$ rollout grid. Each transition substep costs $O(N\log N + K_{\mathrm{rad}}^2 + n_{\mathrm{p}}K_{\mathrm{rad}} + Nn_{\mathrm{p}}^2 + n_{\mathrm{p}}^3)$, including analytic coupling and the tangent solve, and each macro-step uses 16 substeps. The cost statement concerns this implementation, not a hardware-independent speed comparison. The matched learned study equalizes parameters and nominal network structure, but recorded training is roughly $2.0\times$ slower for C than R and $2.2\times$ slower for HC than H. These timings include token construction and online closed-loop recovery, but exclude teacher-forcing feature-cache construction. This accounting supports no learned-efficiency claim. PS64 allocates four times the time substeps of PS16; learned methods also include training cost. The main paper therefore uses matched observations and state controls, not a wall-clock ranking.

Table 3: **Numerical and statistical protocol.** The same field-level unit is used for paired contrasts; horizons are averaged within a field.

Item	Frozen setting	Purpose
Observation/output	$N = 128$ endpoint-free periodic grid	common deployment state and output contract
MS outer truth	1,024-point two-thirds-dealiased ETDRK4; checked at 2,048	separates endpoint effect from truth discretization
MS tail-only truth	2,048-point solver; refined to 4,096 with half time step	checks the isolated tail contrast against truth error
BO truth	1,024-point twofold-zero-padded ETDRK4; 512 substeps per 0.003 macro-step	dense BO reference under its frozen solver
MS temporal controls	PS16 and PS64; 16 and 64 same-grid substeps	tests time refinement without changing the state
B0-D1 ladder	18 fields; $N \in \{64, 128, 256, 512\}$	paired measurement of resolution dependence
MS outer panel	24 fields each in near, transition, radiation-only strata	independent outer replication
MS tail-only panel	256 physical fields; clean plus two 0.5%-RMS noisy views	isolates the high-coordinate factor and its stress boundary
BO outer panel	24 fields in each compatible stratum plus 24 radiation-only fields	72 C0/control rollouts plus rank-zero recovery and spectral boundary controls
Uncertainty	paired field bootstrap; complete trajectories for learned panels	preserves the experimental unit under repeated horizons and retrainings

B Detailed experimental setup

This section records the data generators, observation contracts, truth solvers, controls, split rules, and uncertainty calculations used by the main experiments. The following two sections collect supplementary MS results and the independent BO and learned-operator evidence. Keeping the setup separate from the result expansions makes each main-text claim traceable to a frozen protocol and a corresponding outer record.

B.1 Common data and inference contract

The primary MS and BO rollout endpoints use a 128-point endpoint-free observation and output grid; the stated resolution diagnostics vary N. MS viscosity is sampled from the frozen stratum ranges; the compatible inputs contain one or two separated poles plus visible radiation. BO inputs contain one Poisson travelling wave, low-band radiation, and a small constant background. Every deployment path reads the current coarse field only. Dense decompositions, future states, and high-resolution fields are used only for generation or evaluation.

The primary endpoint is Equation (11). For a paired contrast, the field-level difference is resampled, not the individual horizons. B0-D1 uses a field bootstrap over within-field log-feedback slopes. The MS outer panel uses its frozen 20,000-resample field bootstrap. The C2 three-budget confirmation uses Bonferroni-adjusted 98.333% intervals. These choices preserve the hierarchy of the data and retain the planned meaning of an outer split.

B.2 MS designs

B.2.1 B0–D1 recovery schedule

The B0-D1 fields are generated at 1,024 points and subsampled onto four grids. The audit compares recovered analytic coupling against the sampled-grid product at fixed physical geometry. The matrix-pencil windows are fixed by grid before evaluation: modes 5:24, 13:40, 13:68, and 13:124 at $N = 64, 128, 256, 512$, respectively. Thus the physical field is held fixed while the grid and its predeclared recovery window change together. With $f_{\mathrm{MS}}(v) = \nu\partial_{xx}v + \mathbf{H}\partial_x v - \frac{1}{2}\partial_x(v^2)$, the reported oracle quantity is $\|\mathbf{\Pi}_K\partial_x(((\mathbf{I} - \mathbf{\Pi}_K)p)r)\|_2/\|f_{\mathrm{MS}}(v)\|_2$. The generator-oracle feedback slope uses the known decomposition only for offline measurement and is independent of the recovery window. This is a no-training calculation. The MS outer panel freezes the compatible height ranges, state variants, numerical budget, and bootstrap rule before generating 24 fields in each stratum. The near and transition strata use one or two separated poles, viscosity in the frozen range, and radiation in modes 1–8. The radiation-only panel is a necessity control.

The tail-only panel has two geometry classes (near and transition), two radiation spectra (modes 1–8 and 13–20), and clean/noisy views. T0 masks only $K < |m| \le 2K$ within the analytic pole–radiation

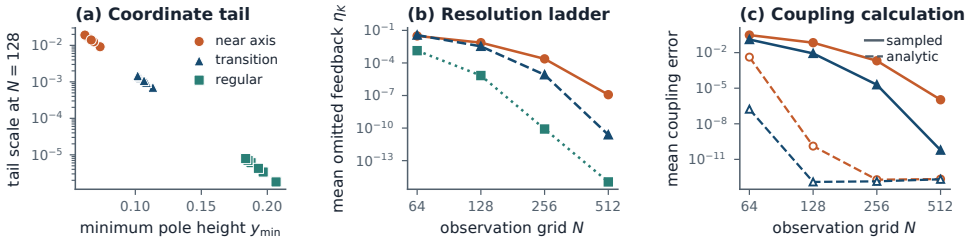

Figure 6: **B0-D1 geometry–resolution map.** Lower pole heights leave larger tails at $N = 128$ (left). On matched physical fields, returned feedback contracts as resolution grows (middle). On the declared near-axis and transition targets, recovered analytic coupling has lower error than the sampled product (right).

convolution. It keeps the raw recovery, decoded state, visible residual, tangent projection, mean gauge, and substep schedule unchanged. The primary clean low-band contrasts are planned paired tests; the remaining cells test the domain over which that narrow causal statement is resolved.

B.3 BO AND LEARNED DESIGNS

The BO exploratory seeds selected an under-resolved height and radiation range before the development and outer registries were created. The formal development registry and 96-field outer registry are disjoint. The BO resolution ladder uses the known generator decomposition offline to measure the designed cross-term magnitude. It is not presented as recovered-coupling accuracy because its analytic term is compared with itself. The raw-field outer rollouts provide that independent operational test.

C1 selects the learned configuration using 256 train and 64 development trajectories, then rebuilds FNO and CNO from three independent initializations for a 64-trajectory outer inventory. C2 creates independent 1,024-trajectory training, 128-trajectory development, and 128-trajectory outer inventories. At 64, 256, and 1,024 training trajectories, FNO is selected anew from its frozen candidates and then rebuilt three times. This design asks how a learned current-field map changes over a finite compatible-data range. It does not give learned methods access to the C0 coordinate or to earlier states.

C ADDITIONAL MS RESULTS AND ABLATIONS

C.1 GEOMETRY AND RESOLUTION MAKE THE TARGET REGIME VISIBLE

Figure 2 presents the paired interaction ladders and recovered coupling errors used in the main-text materiality test. These diagnostic records are separate from the rollout panels below.

Figure 6 reorganizes the B0-D1 records by minimum pole height, grid resolution, and coupling calculation. The left panel connects the analytical condition to the data: lower pole heights leave a longer geometric tail at the deployment grid. The middle panel then follows the same physical fields across grids. The right panel shows that the recovered analytic calculation, rather than the sampled product, tracks the dense projected forcing on the declared under-resolved target cells. This three-part view is why the main paper uses a paired-resolution materiality test before claiming an endpoint effect.

C.2 OUTER MS COMPONENT EFFECTS

The disjoint outer panel is a transition-level replication of the B0-D1 diagnostic. Table 4 reports the near-axis endpoint, where all three component controls are paired with C0 on the same 24 fields. The intervals are entirely below zero. Their relative scales also clarify the architecture: retaining full radiation and assigning forcing to the tangent coordinate have larger effects than the isolated tail mask. C0's smaller tail-only effect is expected because it restores one term within a transition whose other channels remain active.

Table 4: **Independent MS near-axis outer panel.** Five-horizon mean relative L^2 errors over 24 fields. Intervals are for $e_{\mathrm{C0}} - e_{\mathrm{control}}$. All fieldwise comparisons favor C0.

Control	C0 error	Control error	paired 95% interval
PS64	5.35×10^{-5}	1.89×10^{-2}	$[-1.953, -1.807] \times 10^{-2}$
$K16$	5.35×10^{-5}	5.40×10^{-3}	$[-6.314, -4.355] \times 10^{-3}$
NT	5.35×10^{-5}	4.51×10^{-4}	$[-4.763, -3.188] \times 10^{-4}$

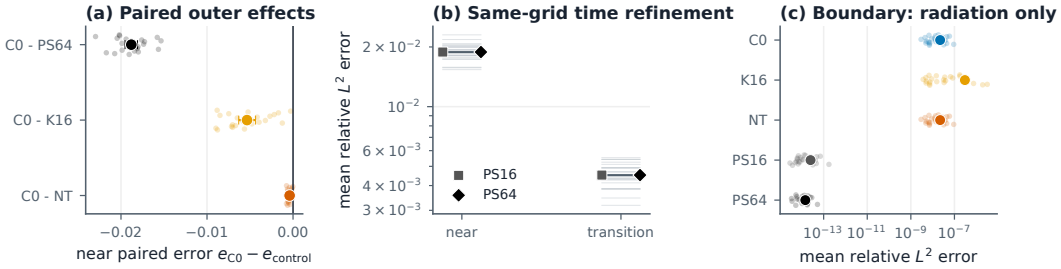

Figure 7: **The full MS transition has a conditional prediction advantage.** Left: paired near-field comparisons favor C0 over PS64, $K16$, and NT. Middle: PS16 and PS64 share a compatible-field time-refinement plateau. Right: the ordering reverses on rank-zero radiation-only fields. Faint points or lines show the 24 paired fields; emphasized marks show the summaries.

Figure 7 complements this table with the time and component controls. The time plot uses the same observed state under PS16 and PS64. Its plateau rules out the particular temporal explanation tested here; it does not make a general claim about all time integrators or time steps.

C.3 Exact observed-band spectral control

The observed-band control uses an independent 72-field registry. FG64 and FG128 retain all observed modes $|k| \leq 63$ and use exact projected quadratic products. The direct coefficient-sum check differs from the padded product by 3.55×10^{-15}. All nine frozen gates pass, including the compatible paired intervals, the radiation-only reversal, and the floor-aware temporal refinement checks.

Table 5: Independent exact-convolution Fourier–Galerkin control. Compatible-row intervals use $\log(e_{\mathrm{C0}}/e_{\mathrm{FG128}})$; the radiation-only row uses the reciprocal ratio.

Stratum	C0	FG128	95% log-ratio interval	Paired wins
Near	6.405×10^{-5}	2.376×10^{-3}	$[-4.006, -3.520]$	C0 24/24
Transition	6.176×10^{-6}	2.709×10^{-4}	$[-4.011, -3.691]$	C0 24/24
Radiation only	1.542×10^{-8}	1.932×10^{-14}	$[-13.796, -12.908]$	FG128 24/24

The saved floor calculation reruns dense reference solutions but no candidate method. Its values require alignment with the original error references, as detailed in Appendix P. Subject to that condition, C0 is below the discrete sample-space floor $e_{\perp}$ on all 48 compatible fields. The reported geometric-mean ratios are 0.080 and 0.074. The same record includes e_{Π}, a descriptive continuum-truncation reference, not a lower bound. Similar FG128 error and e_{Π} do not show that FG128's dynamics are exact. Equation (32) separates the two references.

Radiation-only references have both quantities near numerical precision. FG64 and FG128 differ by about 7×10^{-15} absolute there, despite a large relative ratio; the frozen refinement gate is therefore absolute-floor aware. The compatible-stratum FG64/FG128 agreement is 10^{-10} relative.

Figure 8 gives the original paired errors and coordinate-free reversal; the tables above retain its exact summaries and uncertainty intervals. The reconstructed four-variant diagnostic in Figure 3 compares initialization and sampled output space. It retains continuous-tail aliases and does not separate low-mode dynamics from coordinate decoding (Appendix O.1).

Table 6: **Attainability floor of an observed-band output.** A transition whose state and output lie in the observed band emits a sampled trigonometric polynomial, so its error against the sampled dense truth cannot fall below $e_\perp = \mathrm{dist}(\mathbf{S}_N v, \mathcal{V}_K)/\|\mathbf{S}_N v\|_2$, computed from a regenerated reference solve, without rerunning candidate transitions. The bound assumes a common reference vector; comparisons with the saved method errors retain the reference-replay qualification. Column five counts fields on which all of FG64, FG128, PS64 respect that bound. Column six gives the geometric mean of $e_{C0}/e_\perp$ and the paired 95% interval for $\log(e_{C0}/e_\perp)$; its fieldwise range is 0.012–0.347 (near) and 0.014–0.342 (transition).

Stratum	$e_\perp$	FG128	C0	band-limited $\geq e_\perp$	$e_{C0}/e_\perp$	C0 $< e_\perp$
Near	8.243×10^{-4}	2.376×10^{-3}	6.405×10^{-5}	24/24	$0.0800\ [-2.87, -2.19]$	24/24
Transition	9.555×10^{-5}	2.709×10^{-4}	6.176×10^{-6}	24/24	$0.0737\ [-2.94, -2.27]$	24/24
Radiation only	2.106×10^{-16}	1.932×10^{-14}	1.542×10^{-8}	24/24	–	0/24

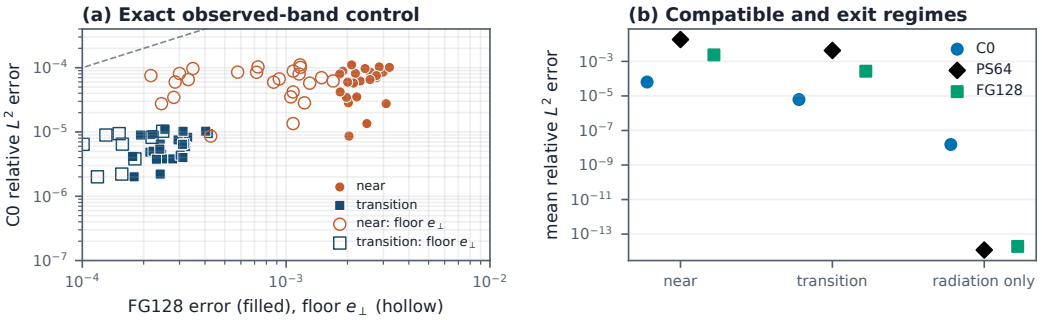

Figure 8: **The MS separation survives exact observed-band convolution.** (a) Each filled point pairs C0 and FG128 on one of 24 fields per compatible stratum; the dashed line marks equal error. Hollow points replace the FG128 error with the saved sample-space floor $e_\perp$, subject to the reference alignment condition in Appendix P. (b) Stratum means compare C0, PS64, and FG128 over the same five forecast horizons. The coordinate-free stratum reverses the ordering. Both panels use logarithmic error scales.

C.4 Complete tail-only panel

Figure 4 gives the main tail-only result, and Table 7 expands it. The two clean low-band cells pass the planned Holm-adjusted tests. The clean overlap cells can have a negative mean log ratio but do not pass the cellwise adjustment, so the main text does not extend the isolated-tail statement to them. All views pass the recovery and finiteness checks; therefore the table separates a tail-specific endpoint contrast from recovery quality.

At $N = 128$, the mean T0–C0 difference on the separate resolution check is 2.42×10^{-6}. It is numerically zero at $N = 256$ (-3.2×10^{-11}, C0 lower on 4/16) and $N = 512$ (10^{-18}). The mean clean endpoint effect, 1.02×10^{-6}, remains far above the 1.58×10^{-12} truth-refinement discrepancy. The 32 radiation-only fields return rank zero and give exact C0–T0 equality. These checks make the resolved exit and coordinate-free exit directly auditable.

D Additional BO and learned results

D.1 BO component interpretation

The BO result is intentionally narrower than an unrestricted transfer claim. It is a raw-field rollout study on a one-wave-plus-radiation family. Its generator-oracle ladder reports a mean within-field log-feedback slope of -0.897 with frozen 95% interval $[-0.909, -0.885]$. That ladder confirms that the chosen BO regime produces the intended resolution trend, but it does not test recovered analytic coupling. The outer C0–SP contrast is the operational measurement. Table 8 reports its two targeted compatible comparisons and the resolved equality; C0 has lower error than SP on 19/24 near

Table 7: **MS tail-only cells.** Ratios are C0/T0. Intervals are paired bootstrap intervals for $\log(e_{C0}/e_{T0})$; the final column gives the planned Holm-adjusted one-sided sign-permutation value.

Geometry	Spectrum	View	C0/T0	95% interval	adjusted p
Near	low	clean	0.97357	$[-0.0550, -0.0251]$	1.60×10^{-4}
Near	low	noisy	0.99999	$[-3.52, 0.617] \times 10^{-5}$	0.335
Near	overlap	clean	0.99555	$[-0.0117, 0.00143]$	0.335
Near	overlap	noisy	0.99987	$[-3.09, 0.470] \times 10^{-4}$	0.335
Transition	low	clean	0.96816	$[-0.0850, -0.0196]$	2.52×10^{-3}
Transition	low	noisy	1.00000	$[-6.39, 23.5] \times 10^{-7}$	0.864
Transition	overlap	clean	0.99836	$[-0.00366, -0.000115]$	0.133
Transition	overlap	noisy	1.00000	$[-8.92, 23.8] \times 10^{-6}$	0.335

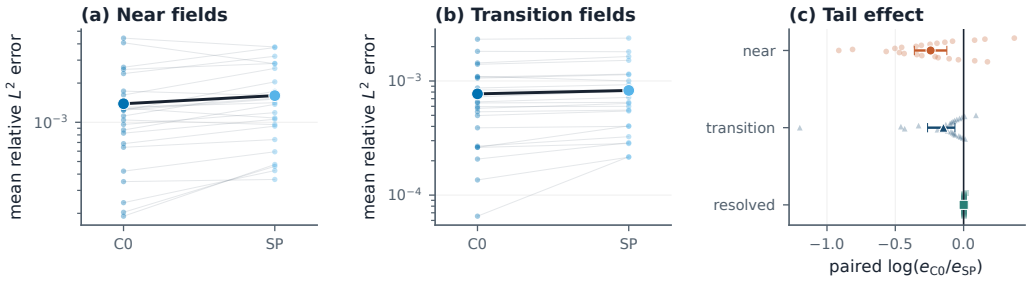

Figure 9: **The BO tail effect disappears when the coordinate tail is resolved.** Panels (a,b) pair C0 with the same-state tail mask SP on each of 24 outer fields; large markers join the field means. Panel (c) displays every paired log-ratio and its field-bootstrap 95% interval. Negative values favor C0. The interval excludes equality in the active strata and contracts to zero after resolution. Full component controls and the offline materiality ladder appear in the appendix.

and 21/24 transition fields, and the resolved stratum splits 13/24. $K16$ and NT have larger effects, consistent with the full closure using both residual bandwidth and geometric assignment.

All 72 compatible BO fields pass rank-one recovery and all 24 radiation-only fields pass rank-zero recovery. The BO result therefore supports a second raw-field realization of the mechanism under a separate decoder and dynamics. It does not merge the BO oracle ladder with the MS recovered-coupling evidence.

D.2 Learned comparisons and their interpretation

The learned panels match forecast observations, but C0 and FNO use different priors. C0 has the equation-specific coordinate and analytic coupling. FNO uses input–output data. The correct conclusion is therefore about fixed-data replacement, not parameter efficiency, universal superiority, or equal-prior comparison. C1 fixes a 256-trajectory budget and uses three independent FNO, CNO, and NOMAD retrainings; the selected FNO has 44,429 parameters, and each C2 arm reselects a 99,473-parameter FNO (Tables 15 and 17). C2 rebuilds and reselects FNO at each data budget before opening its outer inventory.

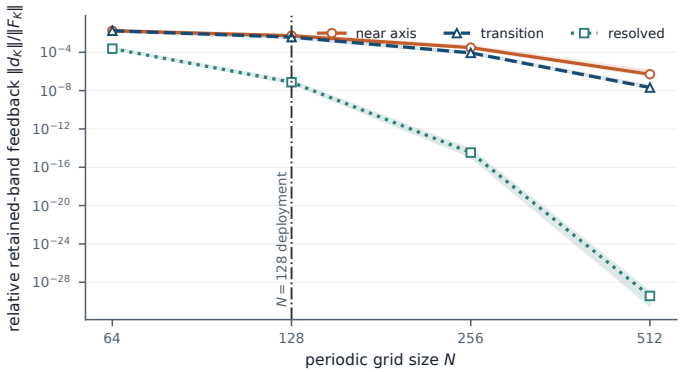

Figure 10: **Offline BO materiality ladder.** The known one-wave decomposition defines the omitted returned-band interaction. The same fields are followed across resolution; the displayed quantities are generator-level materiality diagnostics, not errors of recovered coupling.

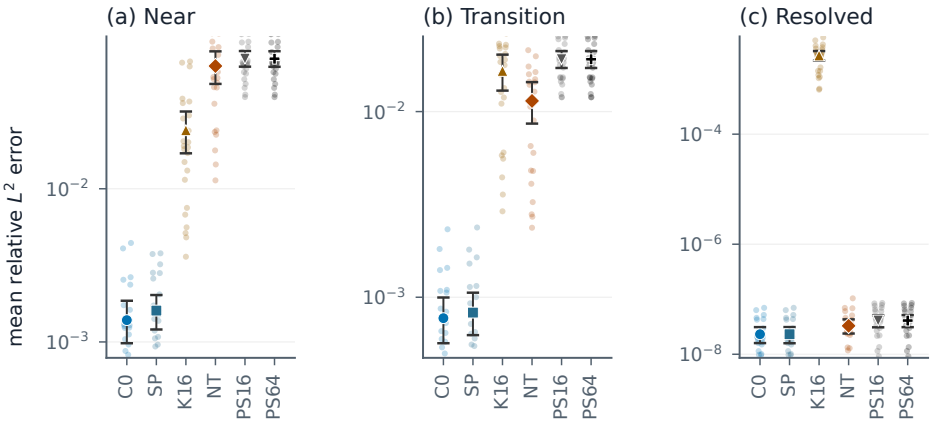

Figure 11: **BO component controls.** Each strip shows all 24 fields on a logarithmic error scale; large markers and bars give field means and 95% bootstrap intervals. SP removes only the unresolved coordinate tail. $K16$ restricts visible radiation, NT suppresses tangent allocation, and PS16/PS64 provide same-grid temporal controls. Separate panel scales preserve the resolved-stratum detail.

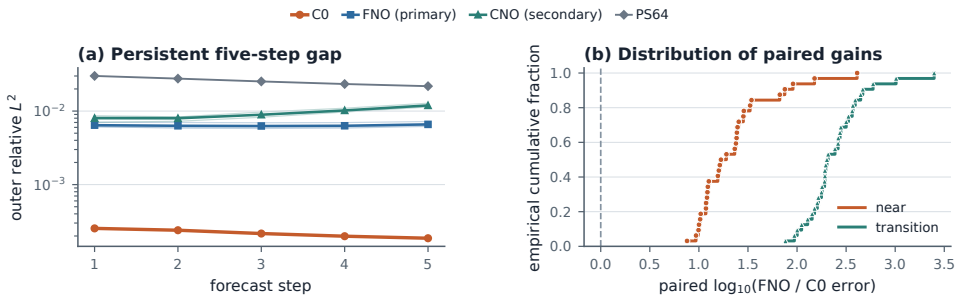

Figure 12: **C1 learned-replacement evidence.** Panel (a) retains every learned retraining curve and its mean across five horizons. Panel (b) is the trajectory-level ECDF of $\log_{10}(e_{\mathrm{FNO}}/e_{\mathrm{C0}})$ after averaging the three FNO retrainings within each trajectory. The zero line denotes equal error; all 64 paired trajectories lie to its right.

Table 8: **BO outer component controls.** Mean five-horizon relative L^2 error over 24 fields per stratum. The interval is the paired field-bootstrap interval for C0–SP.

Stratum	C0	SP	C0–SP interval	$K16$	NT	PS64
Near	1.388×10^{-3}	1.602×10^{-3}	$[-3.99, -0.141]\times10^{-4}$	2.4108×10^{-2}	6.3274×10^{-2}	7.0617×10^{-2}
Transition	7.71×10^{-4}	8.27×10^{-4}	$[-7.92, -3.17]\times10^{-5}$	1.6563×10^{-2}	1.1413×10^{-2}	1.9136×10^{-2}
Resolved	2.31×10^{-8}	2.31×10^{-8}	$[-0.827, 1.42]\times10^{-11}$	2.696×10^{-3}	3.29×10^{-8}	4.10×10^{-8}

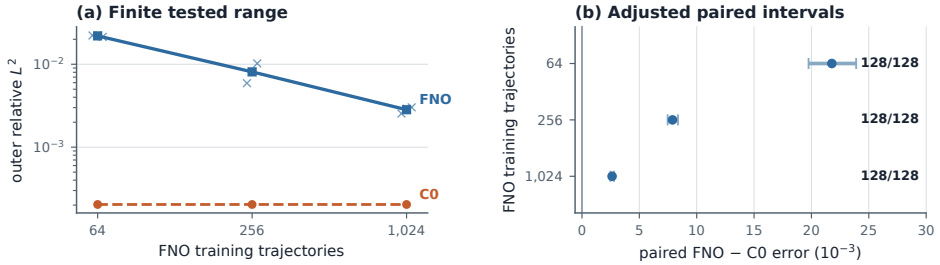

Figure 13: **C2 retains separation across the predeclared data range.** (a) Separately selected FNOs improve as their compatible training set grows from 64 to 1,024 trajectories, while the fixed C0 transition is unchanged. Crosses show the three retained FNO retrainings at each budget; squares show their means. (b) Every Bonferroni-adjusted paired outer interval for FNO−C0 remains positive, and C0 is lower on all 128 trajectories at every tested budget. The supported scope is this finite range. Crossover, equal-prior, equal-capacity, and parameter-efficiency tests require separate designs.

Table 9: **Learned replacement summaries.** C1 uses 64 outer trajectories and C2 uses 128 per budget. C2 FNO errors average three retained retrainings.

Study	train trajectories	C0 mean rel. L^2	FNO mean rel. L^2	paired outcome
C1	256	2.1889×10^{-4}	6.3831×10^{-3}	C0 lower on 64/64
C2	64	2.0354×10^{-4}	2.1981×10^{-2}	C0 lower on 128/128
C2	256	2.0354×10^{-4}	8.0946×10^{-3}	C0 lower on 128/128
C2	1,024	2.0354×10^{-4}	2.8283×10^{-3}	C0 lower on 128/128

FNO improves throughout the tested data range. This monotone response makes the finite-data result more informative than a single selected budget: it establishes a lower bound on the tested range. The C2 protocol intentionally stops at the frozen three budgets; it does not interpolate a crossover or continue training until a desired comparison changes direction.

E MATCHED CURRENT-FIELD COORDINATES AND TEMPORAL HISTORY

This learned study separates a deterministic representation change from the addition of past observations. Six arms share one eight-slot token architecture, the same tensor shapes, and 100,657 real trainable parameters. R receives the current raw field. S receives the recovered coordinate field and visible residual. C adds the returned-interaction field to S. H receives the current field and four past fields, while HS and HC add the corresponding coordinate tokens to H. Coordinate tokens are deterministic functions of the current field. History changes the information set. These frozen arm identifiers are retained in the artifact record; learned arm C is distinct from the deterministic transition C0.

The physical trajectory is the independent unit. The outer registry contains 512 trajectories split across near and transition geometry and low and overlap residual bands. Clean and 0.5% noisy views form paired conditions. Every arm has three finite confirmation runs, and the pooled endpoint gives equal weight to the eight geometry-by-spectrum-by-noise cells.

Table 10: **Matched learned information contracts.** Every arm uses the same eight-slot architecture and learned tensor shapes.

Arm	current field	four past fields	returned token
R	raw	no	no
S	split	no	no
C	split	no	yes
H	raw	yes	no
HS	raw + split	yes	no
HC	raw + split	yes	yes

The pooled C/R contrast passes, but the predeclared learned-robustness gate requires all eight cells to pass. Seven do. In the transition/low-band/clean cell, C/R is 0.9944 and the one-sided 95% upper log bound is 0.0130. The experiment therefore supports a pooled current-field representation benefit for this shared-token FNO, not a uniform learned closure claim. The predeclared H/R and HC/H contrasts do not support a benefit from the tested raw-history input or from adding the complete coordinate-plus-interaction feature set to H.

The secondary arms show that feature utility depends on the surrounding input contract. Without history, decomposition is worse than raw input (S/R $= 1.280$), while the returned token repairs that representation (C/S $= 0.665$). With the tested history branch, HS/H $= 0.808$ but HC/HS $= 1.311$. HS has no frozen primary contrast, so the latter ratios remain descriptive. Together with the failed H/R test, this pattern concerns feature use by one trained architecture; it does not equate current coordinates with past information.

Parameter equality does not imply equal training cost. Recorded training times are 964–977 seconds for R, 1,960–1,962 for S, 1,924–1,933 for C, 930–936 for H, 1,952–1,964 for HS, and 2,088–2,094 for HC. Coordinate extraction during closed-loop updates and token construction are included. Teacher-forcing feature-cache construction precedes the timed training loop and is excluded. The study supports no throughput or wall-clock advantage.

E.1 Nominal parameter matching and active aggregation

The common parameter count is nominal: every arm instantiates the same tensors, but its mask changes the function they implement. R, S, C, H, HS, and HC activate one, two, three, five, seven, and eight slots, respectively. For a query q and active slot set A, the attention weights have the form

$$\alpha_j = \frac{\exp(q^\top k_j/\sqrt{d})}{\sum_{\ell\in A}\exp(q^\top k_\ell/\sqrt{d})}, \qquad j \in A.$$

When $|A| = 1$, $\alpha_j = 1$ independently of the query and key projections; their aggregation gradients vanish. With multiple active slots these pathways can affect the output. The raw-current skip connection is present in every arm. Thus identical tensor shapes and observations support a comparison of the six specified feature and mask combinations, not equality of effective aggregation or an isolated test of the physical meaning of one channel. This observation does not quantify how much of the measured gain is caused by the mask. Equal-active-slot controls and an in-band nonlinear-feature control would address that remaining attribution question; neither was performed.

The pooled C/R result and the descriptive HS/H and HC/HS reversal are consistent with input-dependent feature utility. They do not show that coordinate features substitute for history or help every trained model. The confidence intervals resample physical trajectories after averaging the three confirmation fits, so inference is conditional on those fits rather than over an unrestricted population of training seeds.

The horizon curves are shown in Figure 5; the complete pooled and cellwise contrasts are reported below.

Table 11: **All eight matched C/R cells.** Each geometry–spectrum cell has 128 physical trajectories shared by its clean and noisy conditions. C/R is the ratio of mean errors. The one-sided 95% bound is for the log ratio of cell mean errors. Holm correction applies to permutation p-values computed from paired trajectory log ratios, not to that interval. Three training seeds and repeated noisy views are averaged within a trajectory. The all-cell gate requires every row to pass.

Cell	C/R	95% upper log	Holm p	C wins	Gate
Near/low/clean	0.8544	−0.1284	0.00016	117/128	pass
Near/low/noisy	0.8796	−0.1035	0.00016	113/128	pass
Near/overlap/clean	0.7122	−0.3238	0.00016	128/128	pass
Near/overlap/noisy	0.7555	−0.2669	0.00016	128/128	pass
Transition/low/clean	0.9944	0.0130	0.23636	67/128	fail
Transition/low/noisy	0.9828	−0.0013	0.04288	79/128	pass
Transition/overlap/clean	0.8445	−0.1518	0.00016	124/128	pass
Transition/overlap/noisy	0.8772	−0.1167	0.00016	118/128	pass

E.2 Training objective and input masks

The learned extractor uses two coefficient windows starting at modes 21 and 25, each of length 28, with relative-rank threshold 0.10 and consistency tolerance 0.02. This frozen policy differs from the default C0 window 13:40; it is not an extra learned module or an outer-data selection. Each arm predicts a normalized increment as in Equation (10). Its masks change the active slots, not the parameter shapes. In particular, HS and HC retain the raw current-field slot as well as coordinate features and history. Every arm also uses the same raw-current skip connection. Rejected coordinates set the pole and returned-interaction tokens to zero and retain the residual slot as the raw field for S/C. Fixed type codes, age codes, and recovery confidences remain part of the architecture; this fallback is not a trained R network.

For a training target $\mathbf{v}^{n+1}$, define $\boldsymbol{\delta} = (\mathbf{v}^{n+1} - \mathbf{v}^n)/s_v$ and let $\widehat{\boldsymbol{\delta}} = f_{\theta_a}(\mathbf{T}_a^n, \mathbf{m}_a^n, \widetilde{\nu}, \mathbf{c}_a^n)$. Writing $\mathbf{F}_+$ for the implementation's unnormalized real-input discrete Fourier transform, the per-example loss is

$$\ell = \frac{\|\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}\|_2}{\max(\|\boldsymbol{\delta}\|_2, 10^{-6})} + 0.05\,\frac{\|\mathbf{F}_+(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta})\|_2}{\max(\|\mathbf{F}_+\boldsymbol{\delta}\|_2, 10^{-6})}. \tag{24}$$

The one-sided Fourier norm is the implemented auxiliary loss; it is not claimed to be an independently weighted physical norm. Training uses 15,000 teacher-forced updates followed by 10,000 five-step closed-loop updates, with batch sizes 32 and 8, respectively. The closed-loop loss averages the five per-step losses, with predictions detached between steps rather than back-propagated through the full rollout. AdamW uses weight decay 10^{-4} and gradient clipping at norm 1. The learning-rate candidates are $\{10^{-3}, 3 \times 10^{-4}, 10^{-4}\}$, with development-only selection followed by three fixed-seed confirmation runs. The complete masks and seeds are in `code/matched_coordinate_history/configs/protocol_v1.json`.

E.3 Existing boundary evaluation of the frozen learned models

The secondary boundary record evaluates the same frozen models on 64 resolved and 64 radiation-only trajectories. Each physical trajectory has a clean view and two fixed noisy views. No model is retrained or selected on these fields. The endpoint averages ten horizons and the three confirmation seeds within a trajectory; noisy conditions also average the two noisy views.

C has larger relative and absolute errors than R in every cell, and loses all 64 paired comparisons in each. These are finite predictions, not divergence. The result rules out describing the learned representation as automatically safe once a tail is resolved or a coordinate is absent. A vanishing analytic interaction does not force the weights of C to implement the map learned by R. This boundary concerns the feature-conditioned network, not the algebraic rank-zero equality of deterministic C0 and T0.

The saved summary retains the parent configuration, confirmation-training, outer-analysis, boundary-manifest, and boundary-result hashes. The ratios here are ratios of arithmetic mean errors, not geometric means of paired ratios. They are secondary descriptive boundary results and are not pooled into the primary eight-cell contrast.

Table 12: **Secondary learned-boundary evaluation.** The same frozen C and R models are evaluated without retraining or selection on these fields. Ratios are ratios of arithmetic mean errors; values above one favor R. Errors are averaged over ten horizons and three confirmation seeds within each physical trajectory. The noisy cells also average two fixed noisy views. Each stratum contains 64 trajectories shared by its clean and noisy cells. Wins count paired trajectory-level relative errors.

Stratum	Observation	C/R relative	C/R absolute	C wins
Resolved	Clean	2.407	2.411	0/64
Resolved	Noisy	17.664	17.505	0/64
Radiation only	Clean	9.712	9.336	0/64
Radiation only	Noisy	9.426	9.115	0/64

F Reproduction and audit map

The submission package contains the source and saved records for every regenerable paper figure, together with adapted experiment code for the reported MS, exact observed-band MS, BO, and MS tail-only panels. It separates result generation from package checks. The matched learned display is retained with its hash-pinned summary record; its numerical table is regenerated from that record without retraining.

The asset generator checks every primary JSON digest before reading it. The motivation field is checked from its frozen specification and field statistics to relative tolerance 10^{-9}. A byte check is unsuitable for that decoded sample because PyTorch builds can differ in their final floating-point bits.

The observed-band runner was also replayed on a later PyTorch release. Only one of 72 observation vectors retained byte identity, yet compatible-stratum means agreed to relative error at most 5×10^{-13}. All nine gates and every paired win count were unchanged. Radiation-only means moved at absolute scale 10^{-17}, where a relative comparison is unstable. These differences are recorded as replay precision.

The documented safe study launchers reject destinations inside `data/`. Outer records retain their configuration, source hash, and seed registry. BO and MS use separate registries. Dense generator coordinates remain confined to offline diagnostics. The paper therefore distinguishes source integrity, numerical replay, and scientific evidence.

G Static interface and analytic-core audits

The raw-field encoder and the exact nonlinear interaction need separate checks. The static inverse record below is a five-draw development audit of known Poisson-blurred observations with up to four poles; it is not a test of the current unblurred encoder with at most two poles. The exact-spectral record tests the analytic transition on eight regular-height examples.

The inverse audit supports its declared blurred, fixed-cardinality family. The exact-spectral audit improves on all eight examples and agrees with the refined substep calculation to a ratio of 1.0050. Current raw-field recovery is tested separately in B0–D1 and the rollout panels. The component records neither establish a global recovery guarantee nor substitute for those tests.

H Learned selection and seed accounting

The learned panels separate architecture selection, independent retraining, and outer evaluation. B1b selects an FNO, CNO, and NOMAD configuration using 256 training and 64 development trajectories, with three capacity bands, two learning rates, and three initializations per family. FNO is the sole primary learned comparison in C1 and C2. The other two families are retained as development-selected secondary checks.

C1 rebuilds each selected model from three independent initializations before opening its 64-trajectory outer inventory. The FNO outer mean is stable across these retained retrainings. Reporting the values

Table 13: **Paper-to-artifact map.** Exact package paths identify each display record and the code that consumes or produces it.

Paper evidence	Saved input	Source path
MS resolution mechanism	`data/ms_resolution_records.json`	`code/run_ms_study.py`; asset generator
MS outer transition	`data/ms_outer_records.json`	`code/run_ms_study.py`; asset generator
Exact observed-band MS	`data/ms_full_band_control_v2.json`	`code/run_ms_full_band_control.py`
Observed-band floor	`data/ms_band_limited_floor_v1.json`	`code/ms_band_limited_floor.py`
Reconstructed main diagnostics	`data/main_story_supplements_reconstructed_v1.json`	`code/run_main_story_supplements.py`; `code/generate_main_story_assets.py`
MS tail intervention	`data/ms_tail_only_summary.json`	`code/ms_tail_only/`; asset generator
BO transition and paired effect	`data/bo_records.json`	`code/run_bo_study.py`; `code/bo_model.py`
Matched learned features	`data/matched_coordinate_history/`	`code/matched_coordinate_history/`
C1/C2 replacement exhibits	`data/c1/`; `data/c2/`	retained run records; asset generator
Learned and transfer boundaries	`data/learned_boundary_summary.json`; `data/transfer_boundary_records.json`	asset generator
Exact analytic core	`data/exact_spectral_core/`	asset generator

Table 14: **Separate static prerequisites.** Both audits use independent records and no learned updates. The reference in the first row is a truncated-Fourier inverse; the reference in the second is sampled coupling.

Audit	Fields	Structured quantity	Reference quantity
Known-blur Hardy–Prony inverse	5 development draws	median relative L^2: 3.1402×10^{-6}	5.5608×10^{-2}
Regular-height exact-spectral core	8 independent fields	mean relative L^2: 5.6846×10^{-8}	4.8865×10^{-5}

makes the two sources of variation explicit: retraining seeds describe the learned model, while outer trajectories assess it on held-out physical fields.

For C2, each training budget selects a new FNO from its own development record. The selected architecture has 99,473 active parameters in every arm; the 1,024-trajectory arm selects a smaller initial learning rate. This reselection prevents the data ladder from carrying an unchanged 64-trajectory configuration into a larger-data comparison.

I A DIAGNOSTIC PROTOCOL FOR A NEW COORDINATE DOMAIN

The portable contribution is a testable representation criterion, not a specific pole parameterization. A new nonlinear system needs its own coordinate, decoder, and interaction law. Four gates determine whether a coordinate-closure study is scientifically informative.

1. **Observable recovery.** Recover the coordinate from the forecast-time field and declared parameters, then verify its decoding on a disjoint static audit.

Table 15: **Development-only learned selection.** Scores are equally stratified five-horizon relative L^2 values on the development split.

Family	candidates	selected parameters	LR	development score
FNO	18	44,429	10^{-3}	6.1587×10^{-3}
CNO	18	21,185	10^{-3}	1.1111×10^{-2}
NOMAD	18	88,977	10^{-3}	4.3760×10^{-1}

Table 16: **C1 learned outer retrainings.** Entries are five-horizon relative L^2 averages; FNO is the predeclared primary comparison.

Method	seed 1	seed 2	seed 3	mean
FNO	6.1466×10^{-3}	7.0319×10^{-3}	5.9709×10^{-3}	6.3831×10^{-3}
CNO	9.4952×10^{-3}	8.7204×10^{-3}	1.0015×10^{-2}	9.4103×10^{-3}
NOMAD	4.6001×10^{-1}	4.4800×10^{-1}	5.1206×10^{-1}	4.7336×10^{-1}

2. **Returned-band materiality.** Hold a physical field fixed over resolution and measure the coordinate continuation that re-enters a retained output through its interaction with visible residual content.
3. **Channel attribution.** Preserve the observed input, recovered coordinate, and time budget while changing only the candidate interaction or its assignment between coordinate and residual channels.
4. **Temporal alternative.** Refine the time integrator on the same observed grid. A temporal response indicates that time discretization remains an explanation; a plateau makes the representation hypothesis testable.

The MS sequence completes all four gates. BO completes raw recovery, rollout attribution, and temporal controls on its own one-wave family, while its resolution ladder remains a generator-oracle materiality diagnostic. This distinction protects a future application from turning an illustrative construction into a broad mechanism claim without a recovered-coordinate measurement.

J SEPARATE BO EVIDENCE PATHS

The BO study has two information paths. The first is a raw-field deployment path: each compatible 128-point field is passed through the alias-aware Poisson recovery, C0, and the frozen component controls; a rank-zero field stops at the recovery boundary and is evaluated by PS16/PS64. The second is an offline generator path: the known one-wave decomposition measures the size of the designed omitted cross term over a resolution ladder. Keeping them separate clarifies what each result establishes.

The generator ladder is useful because it verifies that the BO regime has the same resolution signature targeted in MS. Its analytic cross term is compared with itself, so it is not interpreted as a coupling-accuracy measurement. The C0–SP raw-field contrast supplies the operational comparison: it isolates the analytic tail while the other BO controls show the roles of radiation and tangent assignment. The resolved and radiation-only strata supply the corresponding exit conditions. Thus BO corroborates the representation logic without being pooled with the stronger MS direct-coupling record.

K TRANSFER BOUNDARIES

The KdV and Allen–Cahn records test whether static coordinate recovery is sufficient for transfer. It is not. KdV's static coordinate explains its aggregate gain, while learned coordinate updates fail their target-cell or attribution gates. Allen–Cahn recovers the coordinate on all 128 static fields, yet its coordinate–residual forecast has error 0.9297, compared with 0.0332 for FNO. These studies define scope; they are not pooled with the MS or BO evidence.

Table 17: **C2 independently selected FNO configurations.**

Budget	parameters	initial LR	development score	outer retrainings
64	99,473	10^{-3}	2.0945×10^{-2}	3
256	99,473	10^{-3}	6.6860×10^{-3}	3
1,024	99,473	3×10^{-4}	3.0674×10^{-3}	3

Table 18: **BO evidence separation.** The rollout rows use raw inputs; the ladder is an offline materiality diagnostic.

Record	Information available to the calculation	Supported interpretation
Raw-field recovery	current 128-point BO field and frozen normalization	rank-one accepted on $72/72$ compatible inputs; rank-zero on $24/24$ radiation-only inputs
Generator resolution ladder	generator Poisson coordinate and dense decomposition, offline	designed cross-term magnitude contracts with slope -0.897 (95% interval $[-0.909, -0.885]$)
Raw-field C0–SP rollout	current field, shared output grid, shared horizon	tail completion improves the near and transition compatible strata and reaches equality once resolved
PS16/PS64 temporal control	same field, BO equation, and projected state	fourfold time refinement changes the near and transition endpoints by at most 0.212% and 0.092%

L SPLIT-LEVEL LEARNED OUTER REPORTING

C1 uses a fixed 256-trajectory learned-data budget and an independent 64-trajectory outer inventory. Table 20 reports the primary FNO comparison by the two predeclared MS strata. The same direction in both rows rules out an aggregation artifact in which one geometry dominates the outer mean. The uncertainty interval remains a trajectory-paired quantity after averaging the three FNO retrainings within each trajectory.

C2 then regenerates this question with a new 1,024-trajectory training pool, a new development split, and a delayed 128-trajectory outer inventory. Its three FNO arms share neither selected weights nor train-subset normalizers. The consistent paired direction across 64, 256, and 1,024 trajectories supports a finite-range replacement statement. It remains distinct from an equal-prior or asymptotic-capacity comparison because C0 encodes the MS transition while FNO learns it from compatible trajectory pairs.

Table 19: Frozen transfer probes. Each learned study uses separate train, development, and held-out records.

Probe	Frozen result	Disposition
KdV analytic	13/13 positive target records; median $\log_{10}$ gain 0.2686.	Prerequisite passes.
KdV 2-by-2	Aggregate gain 21.5%; target raw FNO is 11.8% lower; M0 wins 1/5.	Target gate fails.
KdV coupled	0.03950 versus raw FNO 0.03141.	Stop rule reached.
KdV tangent	0.022407 versus static coordinate 0.022401.	Learned update adds no gain.
Allen–Cahn	Recovery 128/128; forecast 0.9297 versus FNO 0.0332.	Forecast gate fails.

Table 20: **C1 primary FNO comparison by outer stratum.** Entries are five-horizon means. The interval is the fixed-seed paired bootstrap interval for FNO–C0.

Stratum	C0 rel. L^2	FNO rel. L^2	FNO–C0 interval	C0 wins
Near ($n = 32$)	4.0780×10^{-4}	6.8651×10^{-3}	$[5.7803, 7.1872] \times 10^{-3}$	$32/32$
Transition ($n = 32$)	2.9973×10^{-5}	5.9012×10^{-3}	$[5.4760, 6.2986] \times 10^{-3}$	$32/32$

Table 21: **Claim-to-evidence map.** The manuscript reports these claims separately rather than pooling incompatible evidence.

Claim	Direct evidence	Declared boundary
The coordinate-family continuation restores a returned MS interaction	B0-D1 paired physical-field resolution ladder and analytic-versus-sampled coupling	compatible MS coordinate and visible-residual family
The completed MS transition improves an independent rollout	24-field near and 24-field transition outer controls; PS16/PS64 temporal plateau	128-point grid and short rollout horizon
The result survives a full-band exact-product control	independent FG128 panel with exact products and FG64/FG128 refinement	no claim about every spectral method or unobserved Fourier coefficients
The unresolved tail itself affects the endpoint	cloned-state T0 mask in clean low-band near and transition cells	effect contracts at higher resolution and is unresolved in noisy or overlap cells
The pattern has a second raw-field PDE realization	BO raw recovery, C0–SP, component, temporal, and resolved-tail controls	one-Poisson-wave compatible BO family; ladder remains generator-oracle materiality
Interaction-aware features improve a matched learned model	pooled C/R and C/S contrasts under identical architecture and parameter shapes	seven of eight cells pass; history and learned-boundary failures remain
The fixed MS transition has finite-data learned replacement value	C1 and C2 FNO outer inventories with independent selections and retrainings	tested 64–1,024 trajectory range; distinct physical priors

M CLAIM-TO-EVIDENCE CHECKLIST

Table 21 makes the paper's logical dependencies explicit. Each row names a claim, the record that bears on it, and the condition that limits its scope. This format prevents a strong component result from being used as evidence for a different assertion, such as broad PDE transfer or an equal-prior learned comparison.

Each claim is interpreted only under the information contract in its row. Changing the observation grid, recovery topology, horizon, or learned prior defines a different claim rather than widening an existing one. This separation distinguishes the coordinate-closure mechanism from a broad empirical benchmark.

N FINITE INTERACTION SHELLS AND AN EXACT SINGLE-POLE EXAMPLE

Let $\mathbf{\Pi}_K$ denote continuous Fourier projection on $\mathbb{T}$. Fourier coefficients and the squared L^2 norm use the normalized measure $\mathrm{d}x/(2\pi)$. For $p \in L^2(\mathbb{T})$ and a trigonometric polynomial r supported on $|\ell| \le R$, where $1 \le R \le K$ are integers, define $\Delta_K = -\mathbf{\Pi}_K \partial_x((\mathbf{I} - \mathbf{\Pi}_K)p\, r)$. Projection makes this expression a finite polynomial even when the derivative before projection is interpreted distributionally. Its coefficients satisfy

$$\widehat{\Delta_K}(k) = -ik \sum_{\substack{m+\ell=k \\ |m|>K,\ |\ell|\le R}} \widehat{p}(m)\widehat{r}(\ell), \qquad |k| \le K. \tag{25}$$

For every summand, $|m| \le |k| + |\ell| \le K + R$. Conversely, if $|k| \le K - R$, then $|m| \le K$, so that summand cannot occur. Consequently,

$$\Delta_K = -\mathbf{\Pi}_K \partial_x\big((\mathbf{\Pi}_{K+R} - \mathbf{\Pi}_K)p\, r\big), \qquad \widehat{\Delta_K}(k) = 0 \quad (|k| \le K - R). \tag{26}$$

Thus only the source shell $K < |m| \le K + R$ can affect the output shell $K - R < |k| \le K$.

If $|\widehat{p}(m)| \le c_p e^{-|m| y_{\min}}$ for $c_p \ge 0$ and $y_{\min} > 0$, each surviving coefficient has magnitude at most $K c_p e^{-(K+1)y_{\min}} \sum_{|\ell|\le R} |\widehat{r}(\ell)|$. There are at most $2R$ output coefficients. Parseval's identity yields

$$\|\Delta_K\|_2 \le K\sqrt{2R}\, c_p e^{-(K+1)y_{\min}} \sum_{|\ell|\le R} |\widehat{r}(\ell)|. \tag{27}$$

This is an upper envelope. It allows cancellation and supplies neither a positive lower bound nor a universal slope for a normalized experiment.

An exact MS example realizes the scale in the envelope. Let

$$p(x) = -2\nu \frac{\sin(x-a)}{\cosh y - \cos(x-a)}, \qquad r(x) = \varepsilon \cos(x-\phi), \tag{28}$$

where $\nu, y > 0$, $a, \phi \in \mathbb{T}$, $\varepsilon \in \mathbb{R}$, and $K \geq 1$ is an integer. For $m > 0$, $\widehat{p}(m) = 2i\nu e^{-my-ima}$ and $\widehat{r}(-1) = (\varepsilon/2)e^{i\phi}$. Only $m = K+1, \ell = -1$ and their negative counterparts contribute to Equation (25). Therefore

$$\Delta_K(x) = 2\nu\varepsilon K e^{-(K+1)y} \cos\big(Kx - (K+1)a + \phi\big), \qquad \|\Delta_K\|_2 = \sqrt{2}\nu|\varepsilon| K e^{-(K+1)y}. \tag{29}$$

For $\varepsilon \neq 0$ and integers $K_2 > K_1 \geq 1$, the secant log-slope against $s = (K+1)y$ is exactly

$$\frac{\log \|\Delta_{K_2}\|_2 - \log \|\Delta_{K_1}\|_2}{(K_2 - K_1)y} = -1 + \frac{\log(K_2/K_1)}{(K_2 - K_1)y}. \tag{30}$$

The derivative prefactor explains the finite-resolution correction in this example. Multimode residuals, multiple poles, phase interference, and resolution-dependent normalization preclude identifying this expression with the measured B0–D1 slope without additional assumptions.

O RECONSTRUCTED RETURNED-TERM AND SAMPLED-OUTPUT DIAGNOSTICS

The original `main_story_supplements_v1.json` was absent from the supplied manuscript and code archives. Its supplied runner, `code/run_main_story_supplements.py`, was executed once without changes on the existing resolution and full-band field registries. The reconstruction protocol, source hashes, and expected inventories were recorded before execution in `configs/main_story_reconstruction_v1.json`. The resulting `data/main_story_supplements_reconstructed_v1.json` explicitly records `post_review_reconstruction`; it is not the recovered original file or an independent confirmation. Original frozen records, configurations, failed gates, and learned fits were retained unchanged. No new training, development selection, or field selection was performed.

The reconstruction contains 72 returned-term field–grid records and 72 four-variant forecast records. Its C0-full and FG-raw errors match all 720 corresponding saved horizon errors in the original full-band record exactly in the reconstruction environment. This checks numerical agreement on reused fields, not the provenance of the missing file. The raw output and verification report are retained under `outputs/reconstruction_2026-09-25/`; input hashes and the generated asset manifest link the curated record to the four main-text figures.

The material recovery subset comprises the 24 near and transition records at $N = 64, 128$ with oracle norm above 10^{-12}. Its median relative field error is 7.3×10^{-10}, but the maximum is 1.96%; the identity-line plot of norms alone would conceal this spread. Table 22 reports both by geometry and grid. The full record retains all fields, including the remaining grids and numerical-floor cases. The additional diagnostics are retrospective analyses on previously used registries. Their field-bootstrap intervals are conditional summaries, not new preregistered outer-study gates.

Table 22: **Returned-term reconstruction accuracy.** All 24 material field–grid records are retained. Errors compare recovered and oracle fields, not just their norms. Leakage is the recovered output energy fraction outside the predicted receiving shell; numerical recovery need not preserve exact oracle support.

Geometry	N	Count	Median error	Maximum error	Max. leakage
Near	64	6	5.93×10^{-3}	1.96×10^{-2}	2.35×10^{-4}
Near	128	6	7.20×10^{-11}	4.86×10^{-10}	1.85×10^{-19}
Transition	64	6	1.67×10^{-7}	6.44×10^{-7}	2.18×10^{-13}
Transition	128	6	2.42×10^{-13}	3.71×10^{-13}	3.44×10^{-26}

Table 23: **Reconstructed sampled-output comparison.** Arithmetic mean errors use 24 fields per stratum and five horizons. The ratios below are geometric means of paired field errors with 95% field-bootstrap intervals; they are descriptive, retrospective analyses on the existing registry. Radiation-only FG errors are at the numerical floor.

Geometry	FG-raw	FG-rec	C0-proj	C0-full
Near	2.376×10^{-3}	2.404×10^{-3}	8.290×10^{-4}	6.405×10^{-5}
Transition	2.709×10^{-4}	2.719×10^{-4}	9.599×10^{-5}	6.176×10^{-6}
Radiation Only	1.932×10^{-14}	1.933×10^{-14}	1.542×10^{-8}	1.542×10^{-8}

Geometry	Contrast	Ratio	95% interval	Numerator wins
Near	FG-rec/FG-raw	1.012	$[1.011, 1.013]$	0/24
Near	C0-proj/FG-rec	0.294	$[0.223, 0.384]$	24/24
Near	C0-full/C0-proj	0.079	$[0.056, 0.110]$	24/24
Transition	FG-rec/FG-raw	1.004	$[1.003, 1.004]$	0/24
Transition	C0-proj/FG-rec	0.291	$[0.216, 0.386]$	24/24
Transition	C0-full/C0-proj	0.073	$[0.052, 0.101]$	24/24

O.1 What the common output space does and does not control

Let $\mathbf{S}_N$ sample an endpoint-free even grid and let $\mathbf{P}_N$ project sample vectors onto the non-Nyquist DFT subspace $\mathcal{V}_K$, $K = N/2 - 1$. For the decoded coordinate forecast d, C0-proj is $\mathbf{P}_N \mathbf{S}_N d$. Continuous band truncation instead gives $\mathbf{S}_N \mathbf{\Pi}_K d$. These are generally different:

$$\widehat{\mathbf{S}_N d}_\ell = \sum_{j \in \mathbb{Z}} \widehat{d}_{\ell + jN}, \qquad |\ell| \le K.$$

For example, with $N = 128$, $d(x) = \cos(127x)$ samples exactly as $\cos x$. Thus $\mathbf{P}_N \mathbf{S}_N d = \mathbf{S}_N \cos x$, while $\mathbf{S}_N \mathbf{\Pi}_{63} d = 0$. The non-Nyquist projection removes one sampled direction but leaves aliases of continuous coordinate tails in all retained frequencies. For the smooth analytic fields used here the Fourier alias sum converges.

C0-proj/FG-rec consequently compares complete forecast procedures after matching their sampled output space and improving FG initialization. It does not isolate the vector field, continuous low-mode accuracy, or the unresolved-tail mechanism. C0-full/C0-proj restores only the discrete Nyquist component of the same decoded forecast. The legacy result key `common_output_dynamics` is retained for byte-traceable numerical sections; its interpretation is explicitly corrected in the record's provenance. A comparison of continuous Fourier coefficients against common reference coefficients would address a different question and was not run. The older saved-floor calculation in Appendix P retains its separate reference-alignment qualification.

P Sampled output spaces and a conditional common-subspace diagnostic

Let N be even, $K = N/2 - 1$, and let $\mathcal{T}_K$ be the space of real trigonometric polynomials with frequencies $|k| \le K$. Set $\mathcal{V}_K = \mathbf{S}_N \mathcal{T}_K \subset \mathbb{R}^N$. The real dimension of $\mathcal{V}_K$ is $N - 1$; its orthogonal complement is the single Nyquist direction. Write $\mathbf{P}_N$ for Euclidean orthogonal projection onto $\mathcal{V}_K$. For a nonzero reference sample vector $\mathbf{y} = \mathbf{S}_N v$, define

$$e_\perp = \frac{\|(\mathbf{I} - \mathbf{P}_N)\mathbf{y}\|_2}{\|\mathbf{y}\|_2}, \qquad e_\Pi = \frac{\|\mathbf{S}_N \mathbf{\Pi}_K v - \mathbf{y}\|_2}{\|\mathbf{y}\|_2}. \tag{31}$$

Every $\mathbf{z} \in \mathcal{V}_K$ has relative error at least $e_\perp$. In contrast, the sampled continuum projection $\mathbf{S}_N \mathbf{\Pi}_K v$ need not equal $\mathbf{P}_N \mathbf{y}$: the latter retains aliases from the sampled continuum tail. Orthogonality gives

$$e_\Pi^2 = e_\perp^2 + \frac{\|\mathbf{S}_N \mathbf{\Pi}_K v - \mathbf{P}_N \mathbf{y}\|_2^2}{\|\mathbf{y}\|_2^2}. \tag{32}$$

Thus e_Π is a descriptive truncation error, not an attainability bound. Similar values of a method error and e_Π do not determine the difference between their predictions. C0 and unconstrained grid-output networks need not lie in $\mathcal{V}_K$, so the bound is not a general neural-operator limit.

A useful comparison can be derived without saved forecast arrays. For FG output $\mathbf{z}_{\mathrm{FG}} \in \mathcal{V}_K$ and arbitrary C0 output $\mathbf{z}_{\mathrm{C}}$, let $a = \|\mathbf{z}_{\mathrm{FG}} - \mathbf{y}\|_2/\|\mathbf{y}\|_2$, $c = \|\mathbf{z}_{\mathrm{C}} - \mathbf{y}\|_2/\|\mathbf{y}\|_2$, and $f = e_\perp$. Then

$$\frac{\|\mathbf{P}_N(\mathbf{z}_{\mathrm{FG}} - \mathbf{y})\|_2}{\|\mathbf{y}\|_2} = \sqrt{a^2 - f^2}, \qquad \frac{\|\mathbf{P}_N(\mathbf{z}_{\mathrm{C}} - \mathbf{y})\|_2}{\|\mathbf{y}\|_2} \le c. \tag{33}$$

The first equality follows by decomposing the FG error into orthogonal sample-space components; the second follows from contraction of an orthogonal projection. Consequently $c^2 + f^2 < a^2$ is a sufficient, not necessary, condition for C0 to have lower error in the shared sample subspace.

Table 24: **Conditional shared-sample-band diagnostic.** Counts refer to the sufficient improvement inequality using saved scalar errors. The ratio is the C0 total-error upper bound divided by the FG128 shared-band error, with the square root taken at each horizon before averaging within a field. Ratios are conditional on alignment of the regenerated floor references with the original error references. The records contain no quantitative bound on that discrepancy. The 120 horizons in each row are nested within 24 fields. The diagnostic concerns sampled, potentially aliased truth; it does not establish unaliased continuum accuracy or isolate the tail interaction.

Stratum	Fields	Nested horizons	Geo. mean bound	Maximum bound
Near	24/24	120/120	0.02621	0.06396
Transition	24/24	120/120	0.02397	0.06306

The table applies Equation (33) to the saved five-horizon records. It takes the square root separately at each horizon, averages within each physical field, and only then summarizes fieldwise upper-bound ratios. The compatible rows satisfy the sufficient inequality under the common-reference assumption. Radiation-only rows do not; failure of a sufficient test alone does not determine the reverse ordering. The machine-readable diagnostic retains those rows. No repeated horizon is treated as an independent field.

The numerical interpretation is conditional. The floor file regenerated dense reference solutions from the original configuration and seeds, without rerunning candidate methods. Only one of 72 regenerated initial-observation byte hashes matches the original record. The saved file attributes these differences to floating-point variation, but contains no quantitative bound on future-reference discrepancies. Equation (33) requires a common target and normalization, as well as exact membership of the ideal FG output in $\mathcal{V}_K$; floating-point leakage would also require an error bound for a strict numerical certificate. The table is therefore a post-hoc diagnostic from saved scalars, not an unconditional certificate or independent replay. It compares sampled, potentially aliased truth, not unaliased continuum coefficients, and it does not isolate the tail-specific mechanism.