INTERACTION DEGREE PREDICTS WHEN NONLINEAR CAPACITY IMPROVES NEURAL OPERATORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural-operator capacity is usually chosen from errors observed after training, although the target function itself determines whether nonlinear interactions are needed. We show that interaction degree, measured from reference-normalized response profiles, predicts when added nonlinear capacity will improve accuracy. Whitening removes invertible affine coordinate effects, and homogeneity carries the profile across input radii. Complete homogeneous regression with independently sampled signed rays recovers synthetic polynomial support, agrees across viscous Burgers regimes, and keeps a matched heat operator at degree one. Across 27 Burgers regimes, interaction degree correlates 0.986–0.998 with the gain from a denser polynomial approximation. Across Burgers and Allen–Cahn, 10 of the 12 settings selected by the profile benefit near the training radius; under larger shifts, the mean gain reverses. Interaction degree thus directs nonlinear modules to regimes where they help and warns when distribution shift makes the same capacity harmful.

1 INTRODUCTION

Neural operators share parameters across forcing terms, initial conditions, and physical regimes of partial differential equations (PDEs) (Lu et al., 2021; Li et al., 2021; Kovachki et al., 2023). This sharing gives them their power, but it also hides the structure of the map they learn. When extra capacity helps in one regime and hurts in another, test error alone does not reveal whether the difference comes from nonlinear input interactions, scale, distribution shift, or the geometry of the candidate module. We seek a structural coordinate that can be measured before the architectural intervention is evaluated.

Polynomial and Volterra expansions provide a natural coordinate: degree counts how many input perturbations interact in a homogeneous component (Sandberg, 1982). Raw degree mass, however, changes with the input slice, coordinate scale, expansion point, and operating radius. We therefore ask two separate questions. Can degree be measured reproducibly under a stated reference distribution? If so, when does that measurement predict the value of adding higher-order approximation capacity?

We answer the first question with a reference-normalized profile. Inputs are whitened under the declared slice, and homogeneous components are measured in the output norm. This makes the profile invariant to invertible affine reparameterizations of the same slice. Homogeneity also yields an exact radial transport law for a fixed surrogate, providing a scaling benchmark for distinguishing amplitude effects from changes in dynamics. For nonpolynomial solution maps, the blocks are finite-radius projections rather than assumed Taylor derivatives.

We estimate the blocks in two independent ways. A complete degree-graded regression fits all monomials up to the chosen order. Separately sampled signed rays fit the radial response without sharing features or evaluation points. Agreement between these designs tests whether the measured structure survives the query geometry. A population result then links unnormalized tail energy to nested approximation gain when cross-degree angles and tail coherence are controlled. A counterexample shows why this link cannot by itself select a particular augmentation: two maps can have the same profile and base error but different gains for the same added module.

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.

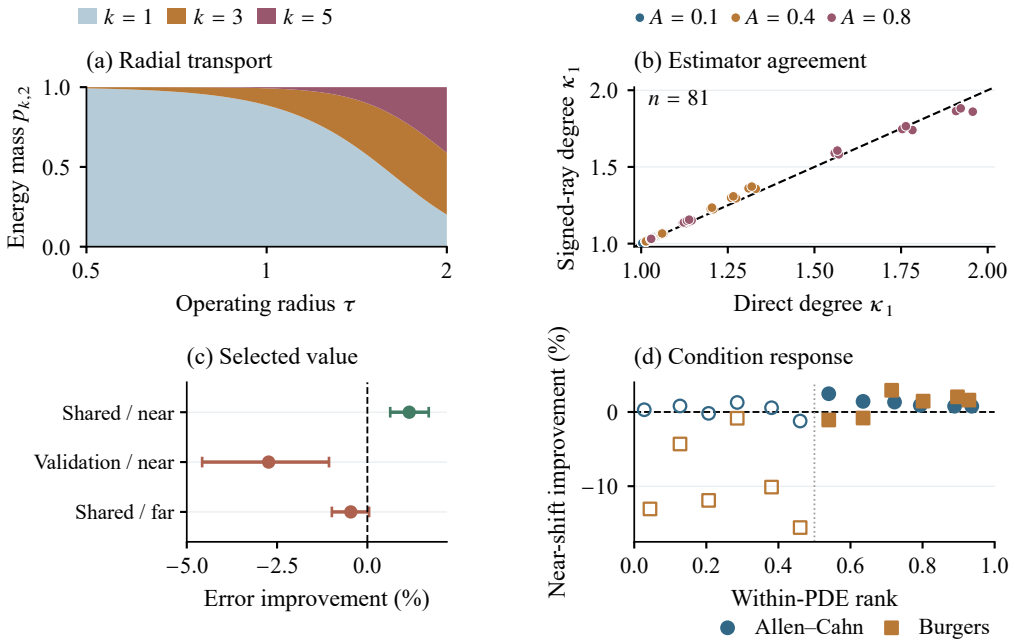


Figure 1: **Interaction degree from measurement to prediction.** (a) Analytic radial law for reference energies $(M_1, M_3, M_5) = (1, 0.12, 0.008)$. (b) Degree estimates for 81 Burgers fits; κ_1 is the root-mean-square-weighted mean degree. (c) Selected improvement (95% replicate-block intervals); shared denotes identical interaction, physical, and outcome-trained selections. (d) Condition response; filled markers denote selection.

The experiments trace a direct path from measurement to prediction. Synthetic maps establish support recovery; Burgers and a matched heat equation separate nonlinear dynamics from radial scaling; repeated coefficient draws connect the profile to the value of denser polynomial spaces. Finally, a Burgers/Allen–Cahn allocation study uses the profile to place a fixed amount of structured residual capacity under distribution shift. It finds a beneficial near-range subset and shows that the same selection need not remain beneficial under a larger shift.

The contribution is a reference-grounded interaction coordinate with explicit transport and identifiability properties, two independent measurement designs, and a matched test of predictive value. Together they show how degree can organize nonlinear approximation value and what residual-alignment information is additionally needed for an architectural choice.

2 Related Work

Neural operators and shared PDE models. DeepONet learns an operator through branch–trunk factorization (Lu et al., 2021), while the Fourier neural operator (FNO) parameterizes translation-equivariant integral layers in frequency space (Li et al., 2021; Kovachki et al., 2021). Later architectures add continuum attention, geometric structure, coordinate deformation, tensorized kernels, or large-scale pretraining (Kissas et al., 2022; Calvello et al., 2025; Hao et al., 2023; Li et al., 2023; Kossaifi et al., 2024; Hao et al., 2024). These methods enlarge or reorganize the operator class. We study the target map from the other direction: measure its interaction structure first, then ask how much approximation value that structure makes available to a candidate class.

Polynomial and Volterra structure. Volterra series organize nonlinear input–output maps into homogeneous multilinear components (Sandberg, 1982). Classical identification already isolates homogeneous responses by applying scaled, sign-alternating copies of a probe and solving Vandermonde or least-squares systems (Boyd et al., 1983). Polynomial kernels can encode order-specific

Volterra/Wiener operators (Franz & Schölkopf, 2006), while regularized estimators impose smooth-decay priors or multiplicative polynomial kernels on their coefficients (Birpoutsoukis et al., 2017; Dalla Libera et al., 2021). These classical results supply the algebra behind our estimators. Our signed-ray design samples independent directions and aggregates radially corrected output norms into reference-law block masses, providing an independent comparison with complete regression.

Orthogonal-polynomial neural operators use boundary-compatible spatial bases for nonperiodic PDEs (Liu et al., 2024), whereas polynomial-chaos methods expand solution operators over uncertain inputs (Sharma et al., 2025). Functional analysis of variance (ANOVA) and Sobol indices instead partition output variance by subsets of input coordinates (Sobol', 2001); total polynomial degree groups all coordinates by radial homogeneity, so the two decompositions answer different questions. Volterra neural networks turn higher-order filters into trainable architectures through cascaded filters or order-specific polynomial-kernel atoms (Roheda et al., 2024; Yun et al., 2026). A complementary algebraic formalism organizes modern layers through product-interaction order (Dong et al., 2026). In these works, order labels structure a learned architecture. Our degree blocks instead belong to a population projection of the target solution map under a pushed-forward reference law. We measure their realized output norms and connect those norms to nested approximation value and candidate alignment. Tensorized operators control large multilinear parameterizations (Jin et al., 2022; Tran et al., 2023; Chou & Chen, 2025). Homogeneous polynomials are symmetric tensors, and sums of powers of linear forms are their classical Waring decompositions (Comon et al., 2008). We use these tools to build a self-contained framework: a declared population object, affine and radial transport laws, independent query-based estimation, and a pre-outcome test of when the statistic predicts value and when candidate alignment becomes decisive. Fourier frequency and polynomial degree are also different notions: spectral bias concerns which spatial frequencies a network learns first (Rahaman et al., 2019), whereas Volterra degree counts how many input perturbations interact.

Conditional allocation. Conditional PDE models incorporate physical parameters directly (Takamoto et al., 2023). Mixture-of-experts architectures allocate or combine specialized capacity: Switch Transformers route examples sparsely, spatial mixtures learn location-dependent weights, and iMOOE aggregates parallel neural-operator experts (Fedus et al., 2022; Deighan et al., 2025; Li et al., 2026). Our allocation experiments ask whether a degree statistic, measured before held-out outcomes, ranks the marginal value of a fixed residual with the same residual capacity. This connects a law-dependent estimand, an independent query design, and a pre-outcome allocation test. The ranking is designed for a declared operating neighborhood; candidate-specific alignment supplies the additional information needed for module selection.

3 Reference-normalized degree and its limits

Volterra degree describes an operator in declared coordinates, not an intrinsic property of a PDE. Let $\mathcal{G}_\theta : \mathcal{X}_\theta \to \mathcal{Y}$ act between normed spaces at physical condition θ. Fix an affine input slice $\iota(c) = a_0 + Ec$, where a_0 is a reference input and E is an injective linear map from a finite-dimensional coefficient space. If a reference coefficient random variable C has mean μ and positive-definite covariance Σ, let $Z = \Sigma^{-1/2}(C - \mu) \sim \nu$ and define the coordinate pullback $F_\theta(z) = \mathcal{G}_\theta(\iota(\mu + \Sigma^{1/2}z))$. For the radii considered, require $\mu + \Sigma^{1/2}\tau \operatorname{supp}\nu$ to remain in the coefficient domain.

A degree profile must attach to a declared polynomial object. For this section let $\mathcal{P}_k$ be the $\mathcal{Y}$-valued homogeneous polynomials of degree k and set $\mathcal{P}_{\le K} = \bigoplus_{k=0}^{K} \mathcal{P}_k$. Assume $\mathcal{Y}$ is Hilbert, $F_\theta \in L^2(\nu; \mathcal{Y})$, $\mathbb{E}_\nu \|Z\|_2^{2K} < \infty$, and no nonzero scalar polynomial of total degree at most K vanishes ν-almost surely. The complete monomial Gram matrix is then finite and positive definite, so the finite-radius population surrogate is

$$Q^\nu_{K,\theta} = \operatorname*{arg\,min}_{Q \in \mathcal{P}_{\le K}} \mathbb{E}_\nu \|F_\theta(Z) - Q(Z)\|_{\mathcal{Y}}^2 = P^\nu_{0,\theta} + \sum_{k=1}^{K} P^\nu_{k,\theta}, \qquad P^\nu_{k,\theta}(\lambda z) = \lambda^k P^\nu_{k,\theta}(z). \quad (1)$$

It is unique in $L^2(\nu; \mathcal{Y})$, and its continuous homogeneous blocks are algebraically unique. We suppress the superscript ν below. The intercept is a nuisance block and is excluded from the profile.

For a fixed surrogate,

$$Q^{\nu}_{K,\theta}(\tau z) - P^{\nu}_{0,\theta} = \sum_{k=1}^{K} \tau^k P_{k,\theta}(z), \qquad \varepsilon^{\nu}_{K,\theta}(\tau, z) = F_\theta(\tau z) - Q^{\nu}_{K,\theta}(\tau z). \tag{2}$$

A Taylor–Volterra interpretation would require differentiability and a separately controlled small-radius remainder; it is not used in our PDE experiments. There P_k denotes a finite-radius projection block, and ε must be measured. Refitting at a new radius produces a different estimand; the transport identity above applies only to the fixed polynomial. For the unit-scale whitened law, define, for $q > 0$,

$$M_k(\theta) = \mathbb{E}_\nu \|P_{k,\theta}(Z)\|_{\mathcal{Y}}^2, \quad p_{k,q}(\tau;\theta) = \frac{\tau^{qk} M_k(\theta)^{q/2}}{\sum_{j=1}^{K} \tau^{qj} M_j(\theta)^{q/2}}, \quad \kappa_q(\tau;\theta) = \sum_k k p_{k,q}(\tau;\theta). \tag{3}$$

We restrict to θ for which $0 < \sum_{j=1}^{K} M_j(\theta) < \infty$ and to radii $\tau > 0$. The choices $q = 1$ and $q = 2$ normalize block root-mean-square (RMS) amplitudes and block energies, respectively; κ_q is the mean degree under the resulting weights. Only $q = 2$ has the direct squared-error interpretation used below; $q = 1$ is retained because the earlier empirical studies froze it before the energy formulation. The general family makes that choice explicit rather than treating its members as interchangeable.

Proposition 1 (Reference normalization and radial transport)**.** *If $c' = Ac + b$ is an invertible affine reparameterization of the same input slice, and the surrogate, expansion point, and reference law are pushed forward, then whitening relates the coordinates by an orthogonal map and leaves every M_k, $p_{k,q}$, and κ_q unchanged. Moreover, for any fixed surrogate and every $q > 0$,*

$$\frac{\mathrm{d}\kappa_q(\tau;\theta)}{\mathrm{d}\log\tau} = q \operatorname{Var}_{p_{\cdot,q}(\tau;\theta)}(k) \geq 0. \tag{4}$$

The proof is in Appendix A.1. The family remains relative to the input slice, expansion point, reference law, output norm, and K. Its $q = 2$ member normalizes component energies; cross-degree inner products mean those masses need not partition the total output energy.

The next two results state precisely what a degree tail can predict and what it necessarily discards.

Proposition 2 (When tail energy predicts nested approximation)**.** *Set $\mathcal{H} = L^2(\nu; \mathcal{Y})$, let Π_j project orthogonally onto $\mathcal{P}_{\leq j}$, and define $\mathcal{E}_j = \inf_{Q \in \mathcal{P}_{\leq j}} \|F_\theta - Q\|_{\mathcal{H}}^2$. For $d < K$, write $T_d = \sum_{k>d} P_{k,\theta}$, $S_d = \sum_{k>d} M_k$, and let n_d be the number of nonzero tail blocks. Let $\gamma_{d,K}$ be the minimum sine of the angle between the tail space $\oplus_{k=d+1}^{K} \mathcal{P}_k$ and $\mathcal{P}_{\leq d}$, and let ρ_d be the maximum absolute normalized inner product between two nonzero blocks in T_d, with $\rho_d = 0$ when $n_d \leq 1$ and $[x]_+ = \max\{x, 0\}$. Then, without an angle assumption,*

$$\mathcal{E}_d - \mathcal{E}_K = \|(I - \Pi_d) T_d\|_{\mathcal{H}}^2, \tag{5}$$

and if $\gamma_{d,K} > 0$, then, with $p_{>d,2} = \sum_{k>d} p_{k,2}$,

$$\gamma_{d,K}^2 [1 - (n_d - 1)\rho_d]_+ S_d \leq \mathcal{E}_d - \mathcal{E}_K \leq [1 + (n_d - 1)\rho_d] S_d, \qquad S_d = p_{>d,2}(1) \sum_{j=1}^{K} M_j. \tag{6}$$

Equality with S_d holds when the tail is orthogonal to the lower-degree space and its active blocks are mutually orthogonal. This is a same-law population squared-loss statement; finite-sample regularization and shifted-law relative error need not inherit its ordering.

Proposition 3 (Degree profile does not identify architecture gain)**.** *For a backbone b and additive class $\mathcal{A}$, define $\Gamma_{\mathcal{A}}(f; b) = \|f - b\|_{\mathcal{H}}^2 - \inf_{a \in \mathcal{A}} \|f - b - a\|_{\mathcal{H}}^2$. If, for $1 \leq k \leq K$, one degree-k subspace contains orthonormal r_1, r_2, take the same backbone $b = 0$, $\mathcal{A} = \operatorname{span}\{r_1\}$, and targets $f_i = r_i$. Both targets have unit base error and the identical pure-k profile for every $q > 0$, yet*

$$\Gamma_{\mathcal{A}}(f_1; 0) = 1, \qquad \Gamma_{\mathcal{A}}(f_2; 0) = 0. \tag{7}$$

The profile discards within-degree orientation, which a restricted architecture needs. Thus architecture gain is not identifiable from degree mass alone.

For a closed linear augmentation space $\mathcal{A}$ and $f \neq b$, the missing variable has an exact form:

$$\Gamma_{\mathcal{A}}(f;b) = \|\Pi_{\mathcal{A}}(f-b)\|_{\mathcal{H}}^2 = a_{\mathcal{A}}(f;b)\|f-b\|_{\mathcal{H}}^2, \qquad a_{\mathcal{A}}(f;b) = \frac{\|\Pi_{\mathcal{A}}(f-b)\|_{\mathcal{H}}^2}{\|f-b\|_{\mathcal{H}}^2}. \tag{8}$$

The identity isolates information absent from degree mass: residual alignment with the augmentation space. For a trained nonlinear class, class geometry and optimization maturity introduce additional variables. Appendix A.2 gives the proofs. The RMS κ_1 and $p_{\geq 3,1}$ used in our experiments are empirical summaries, not theorem-level guarantees of shifted-law gain.

4 From degree measurement to approximation value

4.1 Operator slices and two estimators

For finite-dimensional outputs let $\mathcal{Y} = \mathbb{R}^m$ with $\|y\|_{\mathcal{Y}}^2 = m^{-1}\|y\|_2^2$, the output-coordinate RMS convention used in the experiments. This common factor cancels from every normalized profile. Let $\Phi_k \in \mathbb{R}^{d_k}$ be the empirically whitened complete degree-k monomial block of dimension d_k and $C_k \in \mathbb{R}^{m\times d_k}$. For n training queries (Z_i, Y_i), with operator responses Y_i, the *direct* estimator uses degree-block group lasso (Yuan & Lin, 2006), with Frobenius norm $\|\cdot\|_F$:

$$(\widehat{C}_0,\dots,\widehat{C}_K) \in \operatorname*{arg\,min}_{C_0,\dots,C_K} \left\{ \frac{1}{n}\sum_{i=1}^{n}\|Y_i - \sum_{k=0}^{K} C_k\Phi_k(Z_i)\|_{\mathcal{Y}}^2 + \widehat{\lambda}\sum_{k=1}^{K}\sqrt{d_k}\|C_k\|_F \right\}, \qquad \widehat{P}_k(z) = \widehat{C}_k\Phi_k(z). \tag{9}$$

The penalty is selected on a disjoint validation set. On an independent reference sample of size n_{ref} we report $\widehat{M}_k^{\mathrm{dir}} = n_{\mathrm{ref}}^{-1}\sum_j \|\widehat{P}_k(Z_j^{\mathrm{ref}})\|_{\mathcal{Y}}^2$. Thus, conditional on its split and penalty, $\widehat{Q}_K = \sum_k \widehat{P}_k$ is a finite-sample regularized surrogate. Realized function norms define the reported degree masses, and rank-deficient designs are marked unidentified. Appendix A.1 gives its population target; no finite PDE fit is identified with Taylor derivatives. Every block admits a finite Waring representation, while the primary fits use the complete linear basis; Waring optimization is tested separately under rank misspecification.

The *signed-ray* estimator does not reuse those features or samples. At anchor amplitude a, for independently drawn unit directions v and symmetric nodes t_i, it fits the free-intercept problem

$$(\widehat{\beta}_{a,0}(v),\dots,\widehat{\beta}_{a,K}(v)) \in \operatorname*{arg\,min}_{b_0,\dots,b_K} \sum_i \|F_\theta(at_iv) - \sum_{k=0}^{K} t_i^k b_k\|_{\mathcal{Y}}^2. \tag{10}$$

If the radial reference is $Z_a = aRV$, $R \sim \zeta$, with R independent of the direction law, the estimate from n_v independent directions uses $\widehat{M}_{a,k}^{\mathrm{ray}} = \mathbb{E}_\zeta[R^{2k}]\, n_v^{-1}\sum_{j=1}^{n_v}\|\widehat{\beta}_{a,k}(V_j)\|_{\mathcal{Y}}^2$. For an exact degree-$K$ polynomial and full-rank Vandermonde design, ray least squares returns $\beta_{a,k}(v) = a^k P_k(v)$; the radial correction therefore targets $\mathbb{E}\|P_k(Z_a)\|_{\mathcal{Y}}^2$, the same quantity as the unregularized population direct projection. A finite validation-selected group-lasso fit may be shrunk and need not equal that target exactly. For a nonpolynomial map the two fits are generally distinct finite-radius projections. Agreement then measures stability across query designs; Appendix A.2 gives the exact statements and their relation to a Taylor interpretation.

Within each estimator, its reference energies define the family in Section 3; we identify the member every time it is reported. Primary predictions use the RMS profile κ_1 or its tail mass $p_{\geq 3,1}$, the choice frozen for those experiments. The analytic approximation result motivates the energy profile κ_2; a secondary post-hoc analysis obtains it by squaring the stored block norms without new operator queries. We report the two views separately.

4.2 Predicting value for a fixed residual allocation

Measurement becomes predictive only if it ranks an outcome defined outside the fit used to obtain the degree profile. For a condition θ, let

$$\Delta_{\mathrm{near}}(\theta) = \mathbb{E}\left[\log\frac{L_{\mathrm{base}}}{L_{\mathrm{aug}}} \;\middle|\; \theta,\ \tau \in \mathcal{T}_{\mathrm{near}}\right] \tag{11}$$

be the paired log improvement on the near-shift shell $\mathcal{T}_{\mathrm{near}}$, where L_{base} and L_{aug} are the base and augmented relative ℓ_2 errors. The base is a 60-head degree-1–3 Waring map (960 scalars); the 12-head, two-stage degree-1–3 residual adds 192 scalars and is trained separately for each condition with the same split, two restarts, and 2,000 optimization steps. A validation-selected global scale follows a pointwise residual cap of 10% of the base-output norm. This cap makes the deployed correction a bounded predictor rather than a homogeneous polynomial; here the degree score only ranks its empirical value. Positive Δ_{near} defines benefit. Conditions are ranked by the average within-PDE rank of the direct and signed-ray RMS cubic-plus masses, $p_{\geq 3,1} = \sum_{k\geq 3} p_{k,1}$, and the top m_{act} are activated. Every comparator receives the same quota m_{act}; thus support precision and returned error measure ranking rather than capacity amount. We compare against physical-parameter, labeled-outcome, validation-only, random, and reverse rankings. Conditions outside the selected subset use the base predictor. Selection is condition-level: measuring degree requires reference responses, whereas inference uses the fixed selected predictor.

The same ranking is evaluated on both $\mathcal{T}_{\mathrm{near}}$ and a farther shell $\mathcal{T}_{\mathrm{far}}$. This design reveals the radius over which degree remains predictive: the near-shift result concerns subset ordering, while the farther shell tests whether that ordering survives stronger extrapolation.

5 Experiments

The experiments ask three increasingly demanding questions: whether degree can be recovered, whether it distinguishes nonlinear dynamics, and whether it predicts marginal approximation value under distribution shift. Each stage uses its own data split, and the allocation rule is fixed before the final condition outcomes are evaluated.

Can the profile recover known degree? We generate vector-valued polynomial maps of four input coefficients with pure degrees 1–4, mixed supports $\{1, 3\}$ and $\{2, 4\}$, and the squared readout of a degree-two amplitude. We vary relative output noise from 1% to 10%, training size from 64 to 512, fitted maximum degree, and Waring rank. Highest-active-degree recovery is separated from exact support because small spurious blocks can appear under noise. Complete homogeneous and independently sampled signed-ray estimators use disjoint feature constructions; a separate finite-Waring rank study varies the number of components.

Does the measurement distinguish nonlinear PDE dynamics? For periodic viscous Burgers on $[0, 2\pi)$,

$$\partial_t u + u\,\partial_x u = \nu_{\mathrm{visc}} \partial_{xx} u, \qquad u(0, x) = A \sum_{j=1}^{2} \big(c_{2j-1} \sin jx + c_{2j} \cos jx\big), \tag{12}$$

we cross three amplitudes A, viscosities ν_{visc}, and final times T, use three random seeds, and repeat the solver at three spectral resolutions. The matched heat equation removes $u\partial_x u$ while retaining the same inputs and diffusion. We report direct–ray agreement, truncation error, and the RMS profile κ_1 at reference and shifted radii. Amplitude monotonicity is interpreted through Equation (4); evidence for dynamics comes from estimator agreement, heat separation, and variation in viscosity and time at fixed amplitude.

Does degree predict marginal value under shift? Coefficient-map oracles compare five nested complete polynomial spaces whose feature count grows from 5 to 126, directly measuring the approximation value of additional degree. The initial and repeated analyses share the same 27-regime factorial grid; the repetition uses new coefficient draws on in-distribution (ID) and nearby out-of-distribution (OOD) shells, not new physical regimes, while retaining the original far-OOD failure. We then fix the average direct/ray rank of RMS cubic-plus mass $p_{\geq 3,1}$ before evaluating disjoint physical conditions. In the decisive cross-PDE test, twelve interaction quantiles are drawn from independent 64-condition pools for both Burgers and Allen–Cahn, giving six low- and six high-degree conditions per equation. Every selector activates six conditions per PDE. Near-OOD radii are 1.05–1.25 times the training boundary and far-OOD radii are 1.3–1.5. Labeled-outcome and physical comparators are learned only from closed development matrices; validation selection uses in-distribution error, and the random reference uses 100,000 equal-count subsets.

Uncertainty and decision rules. Paired comparisons use geometric error ratios, so values below one favor the augmentation. Allocation intervals are percentile intervals from 100,000 resamples of ten complete data/model-replicate blocks; they preserve within-block dependence but condition on the fixed 24-condition matrix and candidate-pool selection. Pair counts are descriptive, and Spearman coefficients are complete-matrix point estimates. For the 27-regime replication, a separate sensitivity first averages the three coefficient draws within each regime, then resamples whole regimes and performs leave-one-regime- out deletion. The prespecified joint criterion is evaluated as a whole.

6 RESULTS

6.1 FINITE SAMPLES RECOVER INTERACTION DEGREE

In the noise-free polynomial study, the complete direct-sum estimator recovers exact support in all $21/21$ trials; maximum test and radial-OOD relative errors are 6.8×10^{-16} and 1.23×10^{-15}. Squaring a degree-two amplitude is correctly identified as degree four in all three seeds, confirming that the degree label belongs to the observed map rather than the latent amplitude.

Recovery degrades in interpretable ways. At 1%, 5%, and 10% relative output noise, highest active degree is correct in all $27/27$ direct fits, whereas exact support is correct in $25/27$; the two misses add a small degree-two block to pure degree-four targets at 10% noise. With 2% noise, highest-degree recovery rises from $3/6$ at 64 samples to $6/6$ at 128 samples, and every 64-sample design is correctly flagged as rank deficient. Correct- and over-rank Waring models reach essentially zero error, while rank-one and rank-two fits have median errors 0.634 and 0.280. The signed-ray estimator recovers all $12/12$ clean supports and all $12/12$ supports at 1% noise. Together, the direct and ray studies establish finite-design recovery and show how noise, sample size, and Waring rank determine its reliability.

6.2 BURGERS AND HEAT SEPARATE NONLINEAR DYNAMICS FROM RADIAL SCALING

Across the 81 primary-resolution Burgers fits, direct and signed-ray effective degrees have Spearman correlation 0.990 and median absolute difference 0.00234; the complete resolution study contains 243 fits. At the weakest anchor, direct estimates are 1.0031–1.0034; at the strongest they are 1.909–1.957, with the ray estimator giving 1.860–1.882. All 108 primary-resolution amplitude sequences from independently refitted surrogates are nondecreasing, and the aggregate degree also rises with longer time and lower viscosity. In contrast, all 81 heat-equation controls retain exactly degree one. Resolution changes from 512 to 1024 alter the first 32 solution modes by at most 4.49×10^{-15}.

Figures 1(b) and 2 show the degree estimates across the tested regimes. This study uses the RMS-normalized score κ_1 fixed for these analyses. Equation (4) supplies the scaling benchmark for a transported surrogate; the independently refitted profiles add the dynamical evidence through agreement across estimators, viscosity and time variation at fixed amplitude, and the exact linear control. At the strongest anchor, degree-five truncation error reaches 3.09%–4.27%, marking the resolved boundary of the finite expansion.

6.3 DEGREE PREDICTS NEARBY APPROXIMATION VALUE

Replicated approximation value. Within nested dense coefficient-map truncation oracles, larger RMS-profile degree predicts larger benefit from adding degree-four and degree-five terms. Proposition 2 gives a sufficient bridge for the $q = 2$ energy tail under the same population law. The observed $q = 1$ near-shift association is empirical, and model capacity grows with truncation degree; the experiment therefore measures approximation value rather than parameter efficiency. Across the fixed 27-regime factorial grid, new coefficient draws reproduce eight planned row-level correlations between the direct or ray score and marginal higher-degree value, ranging from 0.986 to 0.998. Averaging the three draws within each physical regime instead gives $\rho = 0.993$–1.000; the lowest lower endpoint among eight regime-bootstrap 95% intervals is 0.957, and every leave-one-regime-out value exceeds 0.992 (Appendix B.2). These intervals quantify stability across the tested regimes; transfer to new regimes remains an open experiment. Every one of 162 regime–replicate records has positive degree-one-to-five and degree-three-to-five gains on both interpolation and near-OOD inputs. The strong-anchor effect reaches a 12.86% near-OOD degree-three-to-five reduction, below

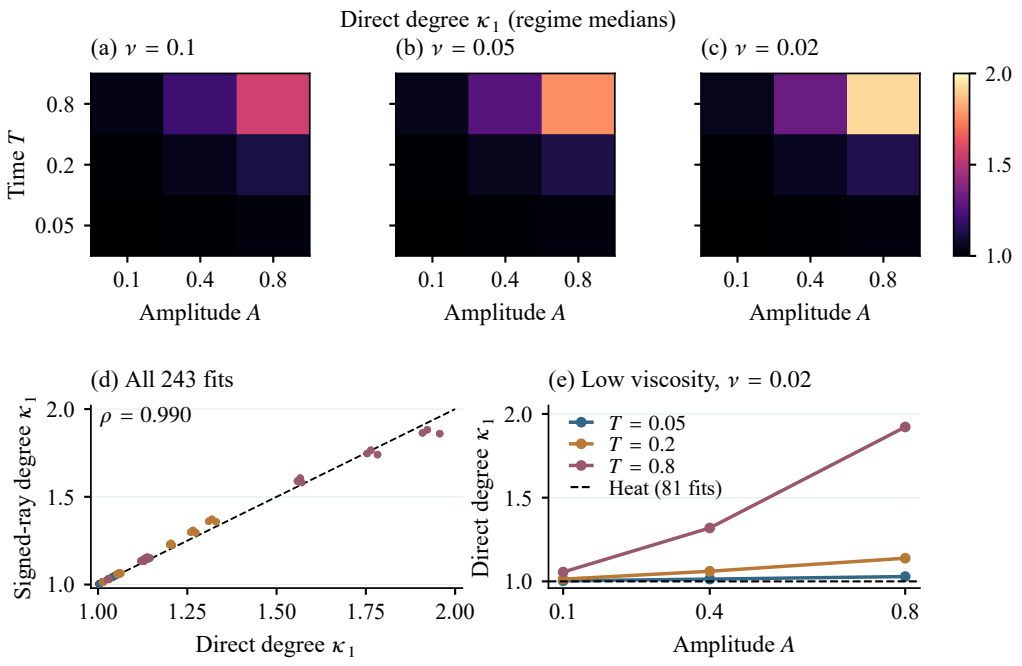


Figure 2: **Interaction degree follows nonlinear Burgers dynamics.** (a–c) Regime medians across seeds and resolutions. (d) All 243 direct and signed-ray estimates agree ($\rho = 0.990$). (e) At $\nu = 0.02$, degree grows with amplitude and time; the dashed baseline shows the common degree-one value of all 81 heat controls.

the chosen 15% threshold. The replicated association is consequently strong across the tested regimes, with magnitude that varies by regime.

Energy-profile sensitivity. As a post-hoc sensitivity analysis, we squared the stored block norms without new operator queries. On the new-draw matrix, the energy tail $p_{\geq 2,2}$ has Spearman correlation 0.986–0.997 with the squared-error reduction $e_1^2 - e_5^2$ across direct and ray estimates on ID and near-OOD inputs. The stricter degree-three-to-five aligned sensitivity comparison is estimator dependent: correlations of $p_{\geq 4,2}$ with $e_3^2 - e_5^2$ are about 0.926 for the direct estimator but 0.996 for rays. Energy weighting therefore reinforces the broad association, while the estimator dependence argues against a universal threshold. These finite-radius experiments and the population theorem answer different questions: the former measures association, while the latter bounds improvement under angle and coherence assumptions. The normalized $q = 2$ tail must be interpreted with the total block energy, estimator, and reference law.

Cross-PDE allocation. A balanced Burgers/Allen–Cahn matrix reveals where that association stops. When each method selects twelve conditions, the direct/ray RMS cubic-plus rank selects $10/12$ beneficial conditions and gives a near-OOD error ratio of 0.988 with interval $[0.983, 0.994]$. The 1.16% point estimate exceeds the chosen 1% practical-effect floor, although its interval includes effects below 1%. The interval resamples ten complete replicate blocks within the fixed 24-condition matrix; $103/120$ paired outcomes improve, as a descriptive count. Replacing $p_{\geq 3,1}$ by the post-hoc $p_{\geq 3,2}$ recoding reconstructs the same 24 conditions and the same twelve selected conditions and hence the same outcomes. This is a sensitivity view of the existing experiment rather than a new $q = 2$ confirmation. Validation-only selection chooses $5/12$ beneficial conditions and is harmful, with ratio 1.027 $[1.011, 1.046]$; exact values and random-subset comparisons are in Appendix B.3.

Operating radius. On a separate far-OOD shell, the four degree–value correlations fall to 0.635–0.758, and at the strongest anchor the degree-five oracle is less accurate than degree one (0.6151 versus 0.5463 relative error). In the allocation study, physical-only and outcome-trained selectors choose exactly the same twelve conditions as the interaction score, so degree supplies structural

meaning to a decision that physical metadata can also recover in this matrix. The pooled area under the receiver operating characteristic curve (AUROC) is 0.782 and Allen–Cahn AUROC is 0.700, both below their chosen floors. On the far-OOD shell, the selected ratio is 1.005 with interval $[0.99953, 1.00983]$: the mean effect reverses and the interval crosses one. The broad preregistered criterion is therefore not satisfied: four of its sixteen criteria are not met, although the narrower near-OOD result remains locally supported. Together these results locate the profile's predictive value near its reference distribution and make larger radial shifts a natural point for remeasurement.

7 Discussion

An interaction-degree profile belongs to a reference experiment. It depends on the chosen input slice, expansion point, probability law, output norm, and truncation order. Whitening removes arbitrary affine coordinates within that experiment, while the radial law predicts how homogeneous components rescale as the input amplitude changes. Together, these two operations turn degree from a raw coefficient summary into a reproducible property of a declared operator slice.

The experiments show that this property can be measured and can predict local approximation value. Direct homogeneous regression and independently sampled signed rays recover synthetic support and agree on Burgers, while the matched heat map remains linear. Across repeated coefficient draws, larger interaction degree consistently marks regimes that gain more from denser polynomial truncation. In the Burgers/Allen–Cahn study, the profile also identifies a near-shift subset that benefits from a structured residual.

The same study locates the operating boundary. Physical metadata and an outcome-trained policy choose the same subset, so the profile contributes a structural explanation for their decision in this matrix. Under the farther shift, its mean benefit disappears. Energy weighting preserves the broad degree–value association but changes the stricter degree-three-to-five comparison across estimators. Degree is therefore most informative near its reference distribution, where its estimand and transport law are explicit.

The theory explains why architecture choice is harder. Under population squared loss, unnormalized tail energy and subspace geometry can bound the gain from a nested approximation space. A particular neural module, however, helps only when its accessible directions align with the remaining residual. Degree measures how much high-order structure is present; it does not specify which architecture can express that structure or whether optimization will find it.

This distinction suggests a concrete design principle for adaptive operators: pair interaction degree with candidate-specific residual alignment. Degree can locate regimes in which nonlinear capacity has value, alignment can choose among modules, and a matched training intervention can test whether the predicted gain is realized. The primary predictive tests use four input coefficients and two one-dimensional nonlinear PDEs. Higher-dimensional fields, longer rollouts, and new equation families provide the next tests of how far the reference profile can be transported.

8 Conclusion

Reference-normalized degree profiles make the interaction structure of an operator map measurable across coordinate systems and operating radii. Two independent estimators recover that structure, and the resulting degree ranks nearby approximation value across synthetic and PDE experiments. Combined with candidate alignment, interaction degree offers a principled route from measuring nonlinear structure to choosing an architecture that can express it.

The measurement and the architecture decision nevertheless remain distinct. A degree profile estimates which interaction orders are present in the target map, whereas a candidate module is useful only when its accessible directions overlap the remaining approximation error. This distinction explains why the profile predicts the value of a denser polynomial space near the reference radius but cannot, by itself, rank arbitrary neural modules or guarantee gains under a larger shift. The natural object for model selection is therefore a pair: a reference-normalized degree profile for the task and an alignment measure for each candidate update space. Declaring the operating radius makes that comparison both interpretable and testable. The benefit of this pairing for selection among trained neural architectures remains to be tested.

Reproducibility statement. The appendix specifies the reference laws, estimators, data splits, PDE solvers, uncertainty procedures, and stopping rule. The supplementary material provides code, configurations, machine-readable reports, and figure data, including unmet criteria and the stopped intervention.

AI use statement. OpenAI Codex was used for coding assistance and language editing. The authors reviewed all AI-assisted content and take full responsibility for the manuscript, implementation, and reported results.

REFERENCES

Georgios Birpoutsoukis, Anna Marconato, John Lataire, and Johan Schoukens. Regularized nonparametric Volterra kernel estimation. *Automatica*, 82:324–327, August 2017. doi: 10.1016/j.automatica.2017.04.014. URL https://doi.org/10.1016/j.automatica.2017.04.014.

Stephen Boyd, Y. S. Tang, and Leon O. Chua. Measuring Volterra kernels. *IEEE Transactions on Circuits and Systems*, 30(8):571–577, August 1983. doi: 10.1109/TCS.1983.1085391. URL https://doi.org/10.1109/TCS.1983.1085391.

Edoardo Calvello, Nikola B. Kovachki, Matthew E. Levine, and Andrew M. Stuart. Continuum attention for neural operators. *Journal of Machine Learning Research*, 26(300):1–52, 2025. URL https://www.jmlr.org/papers/v26/24-0879.html.

Chieh-An Chou and Lu-Hung Chen. Reduced-rank factorized fourier neural operator. In *Proceedings of the 17th Asian Conference on Machine Learning*, volume 304 of *Proceedings of Machine Learning Research*, pp. 702–717, 2025. URL https://proceedings.mlr.press/v304/chou26a.html.

Pierre Comon, Gene Golub, Lek-Heng Lim, and Bernard Mourrain. Symmetric tensors and symmetric tensor rank. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1254–1279, 2008. doi: 10.1137/060661569.

Alberto Dalla Libera, Ruggero Carli, and Gianluigi Pillonetto. Kernel-based methods for Volterra series identification. *Automatica*, 129:109686, July 2021. doi: 10.1016/j.automatica.2021.109686. URL https://doi.org/10.1016/j.automatica.2021.109686.

Dwyer Deighan, Jonas A. Actor, and Ravi G. Patel. Mixture of neural operator experts for learning boundary conditions and model selection. *arXiv preprint arXiv:2502.04562*, 2025. URL https://arxiv.org/abs/2502.04562.

Haonan Dong, Chun-Wun Cheng, and Angelica I. Aviles-Rivero. Product interaction: An algebraic formalism for deep learning architectures. *arXiv preprint arXiv:2602.02573*, 2026. URL https://arxiv.org/abs/2602.02573.

William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39, 2022. URL https://www.jmlr.org/papers/v23/21-0998.html.

Matthias O. Franz and Bernhard Schölkopf. A unifying view of Wiener and Volterra theory and polynomial kernel regression. *Neural Computation*, 18(12):3097–3118, December 2006. doi: 10.1162/neco.2006.18.12.3097. URL https://doi.org/10.1162/neco.2006.18.12.3097.

Beng Yee Gan, Daniel Leykam, and Dimitris G. Angelakis. Fock state-enhanced expressivity of quantum machine learning models. *EPJ Quantum Technology*, 9(1):16, 2022. doi: 10.1140/epjqt/s40507-022-00135-0.

Zhongkai Hao, Zhengyi Wang, Hang Su, Chengyang Ying, Yinpeng Dong, Songming Liu, Ze Cheng, Jian Song, and Jun Zhu. GNOT: A general neural operator transformer for operator learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 12556–12569, 2023. URL https://proceedings.mlr.press/v202/hao23c.html.

Zhongkai Hao, Chang Su, Songming Liu, Julius Berner, Chengyang Ying, Hang Su, Anima Anandkumar, Jian Song, and Jun Zhu. DPOT: Auto-regressive denoising operator transformer for large-scale PDE pre-training. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 17616–17635, 2024. URL https://proceedings.mlr.press/v235/hao24d.html.

Pengzhan Jin, Shuai Meng, and Lu Lu. MIONet: Learning multiple-input operators via tensor product. *SIAM Journal on Scientific Computing*, 44(6):A3490–A3514, 2022. doi: 10.1137/22M1477751.

Nathan Killoran, Thomas R. Bromley, Juan Miguel Arrazola, Maria Schuld, Nicolás Quesada, and Seth Lloyd. Continuous-variable quantum neural networks. *Physical Review Research*, 1(3): 033063, 2019. doi: 10.1103/PhysRevResearch.1.033063.

Georgios Kissas, Jacob H. Seidman, Leonardo Ferreira Guilhoto, Victor M. Preciado, George J. Pappas, and Paris Perdikaris. Learning operators with coupled attention. *Journal of Machine Learning Research*, 23(215):1–63, 2022. URL https://www.jmlr.org/papers/v23/21-1521.html.

Jean Kossaifi, Nikola Kovachki, Kamyar Azizzadenesheli, and Anima Anandkumar. Multi-grid tensorized fourier neural operator for high-resolution PDEs. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=AWiDlO63bH.

Nikola Kovachki, Samuel Lanthaler, and Siddhartha Mishra. On universal approximation and error bounds for Fourier Neural Operators. *Journal of Machine Learning Research*, 22(290):1–76, 2021. URL https://www.jmlr.org/papers/v22/21-0806.html.

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://www.jmlr.org/papers/v24/21-1524.html.

Siyang Li, Yize Chen, Yan Guo, Ming Huang, and Hui Xiong. Towards generalizable PDE dynamics forecasting via physics-guided invariant learning. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=zrQkWSecMR.

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.

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for PDEs on general geometries. *Journal of Machine Learning Research*, 24(388):1–26, 2023. URL https://www.jmlr.org/papers/v24/23-0064.html.

Ziyuan Liu, Haifeng Wang, Hong Zhang, Kaijun Bao, Xu Qian, and Songhe Song. Render unto numerics: Orthogonal polynomial neural operator for PDEs with nonperiodic boundary conditions. *SIAM Journal on Scientific Computing*, 46(4):C323–C348, 2024. doi: 10.1137/23M1556320.

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

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 5301–5310, 2019. URL https://proceedings.mlr.press/v97/rahaman19a.html.

Siddharth Roheda, Hamid Krim, and Bo Jiang. Volterra neural networks (VNNs). *Journal of Machine Learning Research*, 25(182):1–29, 2024. URL https://www.jmlr.org/papers/v25/21-1082.html.

Irwin W. Sandberg. Expansions for nonlinear systems. *Bell System Technical Journal*, 61(2):159–199, 1982. doi: 10.1002/j.1538-7305.1982.tb00331.x.

Himanshu Sharma, Lukáš Novák, and Michael D. Shields. Polynomial chaos expansion for operator learning. *arXiv preprint arXiv:2508.20886*, 2025. URL https://arxiv.org/abs/2508.20886.

Ilya M. Sobol'. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1–3):271–280, 2001. doi: 10.1016/S0378-4754(00)00270-6.

Makoto Takamoto, Francesco Alesiani, and Mathias Niepert. Learning neural PDE solvers with parameter-guided channel attention. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 33448–33467, 2023. URL https://proceedings.mlr.press/v202/takamoto23a.html.

Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized fourier neural operators. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=tmIiMPl4IPa.

Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006. doi: 10.1111/j.1467-9868.2005.00532.x.

Haoyu Yun, Hamid Krim, and Yufang Bao. Learning with volterra neural networks: A system theoretic perspective. *arXiv preprint arXiv:2609.01928*, 2026. URL https://arxiv.org/abs/2609.01928.

A Proofs and algebraic details

A.1 Identifiability and transformation laws

The degree profile concerns realized homogeneous functions, not a factorization of their coefficient tensors. Suppose two degree-at-most-K polynomial representatives agree in $L^2(\nu;\mathcal{Y})$ and let R be their difference. For every $y \in \mathcal{Y}$, the scalar polynomial $z \mapsto \langle R(z), y\rangle_{\mathcal{Y}}$ vanishes ν-almost surely and, by the nondegeneracy assumption in Section 3, vanishes identically. Hence $R \equiv 0$. Writing $R(z) = \sum_{k=0}^{K} R_k(z)$ and restricting $R(tz)$ to each line, uniqueness of univariate coefficients gives $R_k \equiv 0$ for every k. Thus the homogeneous functions are unique even though a Waring or symmetric-CP factorization inside one block need not be.

For a nonpolynomial operator, the unique fitted blocks are Taylor–Volterra components only for an explicit Taylor truncation under sufficient smoothness. Our PDE blocks are finite-radius regressions, whose interpretation depends on the slice, fitting law, expansion point, output metric, regularization, and truncation order.

Population target and finite-sample estimation. Let $\Psi_{\le K}$ collect a fixed monomial basis of the complete degree-at-most-K polynomial space, and suppose the output is finite dimensional. Under a fitting law ρ, finite second moments, and a positive-definite Gram matrix $G_\rho = \mathbb{E}_\rho[\Psi_{\le K}(Z)\Psi_{\le K}(Z)^\top]$, the population least-squares coefficient matrix is unique:

$$B^*_\rho = G_\rho^{-1}\mathbb{E}_\rho[\Psi_{\le K}(Z)F_\theta(Z)^\top]. \tag{13}$$

For i.i.d. data, finite fourth feature moments, and penalties $\lambda_n \to 0$, the fixed-basis estimator

$$\frac{1}{n}\sum_{i=1}^{n}\|F_\theta(Z_i) - B^\top\Psi_{\le K}(Z_i)\|^2_{\mathcal{Y}} + \lambda_n\sum_{k=1}^{K}\sqrt{d_k}\|B_k\|_F \tag{14}$$

has coefficients and fitted function converging in probability to B^*_ρ and its projection. The empirical Gram and cross-moment converge; positive definiteness gives local uniform strong convexity, and the vanishing penalty leaves the limiting minimizer unchanged. This applies to a fixed raw or population-whitened basis, not our empirical whitening with validation-selected penalties. Mapping coefficients back through the invertible empirical whitener preserves the fitted homogeneous function. Thus coefficient matrices depend on within-block coordinates, whereas realized functions and their reference norms M_k do not.

Coordinate invariance. Let one coefficient chart of the fixed input slice have reference mean μ and positive-definite covariance Σ. Consider an invertible affine reparameterization $c' = Ac + b$ of the same slice. Push the expansion point and reference law forward, so

$$\mu' = A\mu + b, \qquad \Sigma' = A\Sigma A^\top. \tag{15}$$

The two whitened coordinates satisfy

$$z' = \Sigma'^{-1/2}(c' - \mu') = Qz, \qquad Q = \Sigma'^{-1/2} A \Sigma^{1/2}. \tag{16}$$

Because $QQ^\top = I$, Q is orthogonal. Changing variables preserves the projection objective, so its unique minimizer pushes forward. If P_k is the degree-k component in the first chart, the same operator in the second chart has $P'_k(z') = P_k(Q^\top z')$. Pushing the whitened reference law ν through Q therefore gives

$$\mathbb{E}_{z' \sim Q_\# \nu} \|P'_k(z')\|_{\mathcal{Y}}^2 = \mathbb{E}_{z \sim \nu} \|P_k(z)\|_{\mathcal{Y}}^2 = M_k. \tag{17}$$

Every $p_{k,q}(\tau)$ and $\kappa_q(\tau)$ is consequently unchanged. The same argument covers any two valid whitening square roots, which differ only by an orthogonal map. This theorem does not compare different input slices, reference measures, expansion points, output metrics, or K.

Radial transport. Set $s = \log \tau$ and, for any $q > 0$,

$$p_{k,q}(s) = \frac{e^{qks} M_k^{q/2}}{\sum_j e^{qjs} M_j^{q/2}}. \tag{18}$$

Direct differentiation yields

$$\frac{\mathrm{d}p_{k,q}}{\mathrm{d}s} = q p_{k,q}(k - \kappa_q), \qquad \kappa_q = \sum_k k p_{k,q}. \tag{19}$$

Hence

$$\frac{\mathrm{d}\kappa_q}{\mathrm{d}\log\tau} = q \sum_k k p_{k,q}(k - \kappa_q) \tag{20}$$

$$= q\left(\mathbb{E}_{p_{\cdot,q}}[k^2] - \mathbb{E}_{p_{\cdot,q}}[k]^2\right) = q \operatorname{Var}_{p_{\cdot,q}}(k) \geq 0. \tag{21}$$

The derivative is strictly positive exactly when at least two distinct degrees have positive mass. In particular, the RMS profile κ_1 has coefficient one and the energy profile κ_2 has coefficient two. They are members of the same reference-normalized family but remain different numerical estimands.

The $q = 2$ profile normalizes component energies. In general, $\mathbb{E}\|\sum_k P_k(Z)\|^2$ also contains cross-degree inner products, so $p_{k,2}$ need not partition total output energy. Pairwise orthogonality under the reference law is sufficient for such a partition, although cancellation among cross terms can also make the totals agree.

A.2 Approximation value and independent recovery

Let $Q_K^\nu = \Pi_K F_\theta$ be the population projection in Equation (1). Since $F_\theta - Q_K^\nu$ is orthogonal to $\mathcal{P}_{\leq K}$, Pythagoras gives, for every $Q \in \mathcal{P}_{\leq d}$,

$$\|F_\theta - Q\|_{\mathcal{H}}^2 = \|F_\theta - Q_K^\nu\|_{\mathcal{H}}^2 + \|Q_K^\nu - Q\|_{\mathcal{H}}^2. \tag{22}$$

Writing $Q_K^\nu = L_d + T_d$ with $L_d = \sum_{k \leq d} P_k \in \mathcal{P}_{\leq d}$ and minimizing over Q yields

$$\mathcal{E}_d - \mathcal{E}_K = \operatorname{dist}(T_d, \mathcal{P}_{\leq d})^2 = \|(I - \Pi_d) T_d\|_{\mathcal{H}}^2. \tag{23}$$

Precisely, the subspace-angle constant in Proposition 2 is

$$\gamma_{d,K} = \inf_{0 \neq T \in \oplus_{k=d+1}^K \mathcal{P}_k} \frac{\|(I - \Pi_d)T\|_{\mathcal{H}}}{\|T\|_{\mathcal{H}}}, \tag{24}$$

and the tail coherence is zero for at most one active block and otherwise

$$\rho_d = \max_{\substack{d < k < \ell \leq K \\ M_k M_\ell > 0}} \frac{|\langle P_k, P_\ell \rangle_{\mathcal{H}}|}{\sqrt{M_k M_\ell}}. \tag{25}$$

The angle definition bounds the exact distance between $\gamma_{d,K}^2\|T_d\|_{\mathcal{H}}^2$ and $\|T_d\|_{\mathcal{H}}^2$. Moreover,

$$\|T_d\|_{\mathcal{H}}^2 = S_d + 2\sum_{d<k<\ell\le K}\langle P_k, P_\ell\rangle_{\mathcal{H}}. \tag{26}$$

Applying the coherence definition and $2\sqrt{M_kM_\ell} \le M_k + M_\ell$ proves Equation (6).

For completeness, set $A_1 = \sum_{k=1}^K \sqrt{M_k}$. If $n_d \ge 1$, Cauchy–Schwarz gives

$$\frac{A_1^2 p_{>d,1}(1)^2}{n_d} \le S_d \le A_1^2 p_{>d,1}(1)^2. \tag{27}$$

Together with Equation (6), this envelopes the RMS tail only when A_1, coherence, angle, and active-tail count are controlled. The empirical κ_1 is a degree mean rather than this tail, and the $p_{\ge 3,1}$ score does not exactly match every tested truncation boundary; its cross-condition association is therefore not a theorem consequence. If $n_d = 0$, both S_d and $\mathcal{E}_d - \mathcal{E}_K$ are zero.

To prove Proposition 3 constructively, fix $1 \le k \le K$ and take a rotation-invariant reference law with positive density on an annulus in $\mathbb{R}^2$. Normalized real and imaginary parts of $(z_1 + \mathrm{i}z_2)^k$ are orthonormal homogeneous degree-k polynomials; call them r_1, r_2. With $b = 0$ and $\mathcal{A} = \operatorname{span}\{r_1\}$, both targets have unit base error and the same pure-degree profile. Orthogonal projection onto $\mathcal{A}$ removes r_1 exactly and leaves r_2 unchanged, giving gains one and zero. The counterexample already holds with oracle optimization. Finite-sample and optimization analyses require additional information and cannot restore identification from degree mass alone.

For the constructive identity in Equation (8), let $r = f - b$ and let $\Pi_{\mathcal{A}}$ be the orthogonal projection onto the closed linear subspace $\mathcal{A}$. Pythagoras gives $\inf_{a\in\mathcal{A}}\|r-a\|_{\mathcal{H}}^2 = \|r\|_{\mathcal{H}}^2 - \|\Pi_{\mathcal{A}} r\|_{\mathcal{H}}^2$. Subtracting from the base error proves the identity. It isolates the missing alignment information rather than inferring it from degree mass.

Independent signed-ray recovery. For a direction v, stack the signed nodes in a Vandermonde matrix $\mathbf{A}_{ik} = t_i^k$, $k = 0, \dots, K$; the implementation fits the intercept rather than fixing it from a separate evaluation. If $F_\theta(z) = \sum_{k=0}^K P_k(z)$ exactly, columnwise least squares at anchor a returns $\beta_{a,k}(v) = a^k P_k(v)$ whenever $\mathbf{A}$ has full column rank. If the direct reference factors as $Z_a = aRV$, with $R \sim \zeta$ independent of $V \sim \sigma$, then

$$\mathbb{E}_\zeta[R^{2k}]\,\mathbb{E}_\sigma\|\beta_{a,k}(V)\|_{\mathcal{Y}}^2 = \mathbb{E}\|P_k(aRV)\|_{\mathcal{Y}}^2. \tag{28}$$

Thus the radial-moment correction in the ray estimator targets the direct reference mass in the exact-polynomial case. With stacked observation error and truncation remainder e_v,

$$\|\widehat{\boldsymbol{\beta}}_a(v) - \boldsymbol{\beta}_a(v)\|_{\ell_2(\mathcal{Y})} \le \|\mathbf{A}^\dagger\|_2 \|e_v\|_{\ell_2(\mathcal{Y})}. \tag{29}$$

Signed nodes are not required for algebraic identifiability, but they separate odd and even responses and make parity violations visible. Here $\mathbf{A}^\dagger$ is the Moore–Penrose inverse and the matrix norm is its induced operator norm. Directions and query points used by this estimator are independent of the direct homogeneous fit. For nonpolynomial F_θ, however, the ray coefficients describe the node-law projection of $t \mapsto F_\theta(atv)$ and need not restrict the direct population projection. Agreement is then empirical cross-design stability.

A.3 Waring representations and nonlocal realization

Continuous-variable models provide nonlinear bosonic feature maps (Killoran et al., 2019), and fixed photon number can expand representational capacity (Gan et al., 2022). We use this connection only to state a classical parameterization identity. It is not needed by either primary estimator.

Let $a_j^\dagger$ create a boson in mode j, and let $|0\rangle$ be the vacuum. For $c \in \mathbb{R}^M$, let $\rho = \|c\|_2$, $\widetilde{c} = c/\rho$ when $c \ne 0$, and

$$b^\dagger(\widetilde{c}) = \sum_{j=1}^M \widetilde{c}_j a_j^\dagger, \qquad |\psi_k(c)\rangle = \frac{(b^\dagger(\widetilde{c}))^k}{\sqrt{k!}}|0\rangle. \tag{30}$$

For an occupation vector n with $|n| = \sum_j n_j = k$, let $n! = \prod_j n_j!$ and $|n\rangle$ denote its normalized occupation state. The multinomial expansion gives

$$|\psi_k(c)\rangle = \sum_{|n|=k} \sqrt{\frac{k!}{n!}}\, \widetilde{c}^{\,n} |n\rangle. \tag{31}$$

Define $\rho^k \langle \eta_k | \psi_k(c) \rangle$ at $c = 0$ by its continuous extension, zero; the resulting function is a homogeneous polynomial of degree k. Choosing the condensate kernel

$$|\eta_k(w)\rangle = \frac{(\sum_j w_j a_j^\dagger)^k}{\sqrt{k!}} |0\rangle \tag{32}$$

for $w \in \mathbb{R}^M$ gives exactly $\langle w, c \rangle^k$. Real symmetric tensors are spanned by rank-one symmetric powers, so finite linear combinations cover every homogeneous block represented in the complete occupation basis and are therefore classical Waring forms for the homogeneous blocks in Equation (9). This is a parameterization identity, not the primary fitting algorithm or a claim of quantum advantage. Squared probability readout has degree $2k$, so the degree-two amplitude in the synthetic control has a degree-four probability map.

Nonlocal realization. A fixed real-even spectral bank $q[u]$ supplies nonlocal linear coordinates on a periodic grid. Two bias-free, context-conditioned Waring stages use powers $\{1, 3\}$, with a B-channel skip between them:

$$h_\psi = q[u] + \mathcal{W}^\theta_{1,\psi}(q[u]), \qquad R_\psi(u, \theta) = \mathcal{W}^\theta_{2,\psi}(h_\psi). \tag{33}$$

Proposition 4 (Spectral Waring–Volterra correspondence)**.** *For fixed θ and finite heads, R_ψ has unique homogeneous field components only in $\{1, 3, 5, 7, 9\}$. Each degree-k component is a translation-equivariant Volterra polynomial with a finite symmetric separable kernel drawn from the spectral-bank span. It is nonlocal when the aggregate kernel has off-origin support. Some components may vanish, so nine is a bound, not an assertion of active degree nine; identity multipliers collapse the construction to a pointwise odd polynomial.*

A fixed global residual scale preserves its relative degree profile unless it is zero. These facts certify structural access, not performance.

On a periodic grid, let m_b be fixed real-even multipliers and define

$$q_b[u](x) = \mathcal{F}^{-1}[m_b \widehat{u}](x) = (\chi_b * u)(x), \qquad b = 1, \ldots, B. \tag{34}$$

For spatially constant context θ, set $D_\ell(\theta) = \mathrm{diag}(\mathbf{1} + \tanh(G_\ell \theta))$ and

$$\mathcal{W}^\theta_{\ell,\psi}(v)(x) = \sum_{s \in \{1,3\}} \sum_{h=1}^{H_\ell} a_{\ell s h} \langle \bar{w}_{\ell h}, D_\ell(\theta) v(x) \rangle^s, \tag{35}$$

where $\bar{w}_{\ell h} \in \mathbb{R}^B$, $a_{\ell s h} \in \mathbb{R}^{C_\ell}$, $C_1 = B$, and $C_2 = 1$. The residual in Equation (33) uses a B-channel skip between these stages.

The implementation uses $\bar{w} = w / \max\{\|w\|_2, 10^{-12}\}$; the normalization affects scale but not homogeneous degree. The field experiment has $d = 2$, $N = 32$, $B = 8$, three physical context coordinates, and 64 heads in each Waring stage. The mask Gram matrix is conditioned using a fixed pre-normalization spectral estimate, which is not asserted to equal either the post-normalization field covariance or the residual profile law introduced below. The first stage maps $\mathbb{R}^B$ to $\mathbb{R}^B$, whereas the second maps $\mathbb{R}^B$ to $\mathbb{R}$.

Fix the physical context. Every spectral coordinate $q_b[u]$ is a linear convolution of u. The first Waring stage therefore has homogeneous field degrees one and three, and its skip connection yields $h = h_1 + h_3$. The second stage's linear branch preserves degrees one and three. Expanding its cubic branch by multilinearity gives degrees three, five, seven, and nine; no even degree or constant term can appear.

At a fixed spatial location, each resulting component is a homogeneous polynomial in the finite vector of spectral coordinates. Polarization makes it the diagonal of a unique symmetric multilinear

form. A finite Waring expansion expresses that form as a finite sum of rank-one symmetric tensors. Substituting $q_b[u](x) = \sum_y \chi_b(x-y)u(y)$ into each factor gives a kernel of the form

$$\mathcal{K}_{k,\psi,\theta}(x-y_1,\dots,x-y_k) = \sum_{m=1}^{J_k} \gamma_{km}(\psi,\theta) \prod_{j=1}^{k} \varphi_{km,\psi,\theta}(x-y_j), \tag{36}$$

with every $\varphi_{km,\psi,\theta}$ in the span of the fixed spectral kernels. This proves finite symmetric CP rank. Translating u cyclically translates each convolution response; all remaining maps use spatially constant coefficients and act pointwise on those responses, proving translation equivariance. Replacing every multiplier by the identity collapses the input to copies of $u(x)$ and gives a pointwise odd polynomial, establishing the role of the spectral bank in making nonlocal dependence available.

For the residual alone, let $U \sim \nu_R$ be a declared field law satisfying $\mathbb{E}_{\nu_R}\|U\|^{18}_{\mathcal{X}_\theta} < \infty$, and write $M_k^{(R)}(\psi,\theta) = \mathbb{E}_{\nu_R}\|R^\theta_{\psi,k}(U)\|^2_{\mathcal{Y}}$. Assume $0 < \sum_{k\in\{1,3,5,7,9\}} M_k^{(R)}(\psi,\theta) < \infty$. Multiplication by one input-independent scalar $0 < \alpha \le 1$ sends every $M_k^{(R)}(\psi,\theta)$ to $\alpha^2 M_k^{(R)}(\psi,\theta)$, so the common factor α^q cancels from every normalized q-profile. At $\alpha = 0$ the residual and all $M_k^{(R)}(\psi,\theta)$ vanish, so that profile is undefined. A sample-dependent clipping multiplier generally destroys homogeneity and does not have this property.

B Experimental details

B.1 Synthetic and PDE measurement protocols

The noise-free study uses four input coordinates, twelve output coordinates, and seven target supports: pure degrees one through four, mixtures $\{1,3\}$ and $\{2,4\}$, and the squared readout of a degree-two amplitude. Three deterministic seeds give 21 trials. Training radii lie in $[0.25, 1.25]$ and radial-OOD radii in $[1.5, 2.0]$.

The robustness study tests pure degree four and the two mixed supports at 1%, 5%, and 10% RMS-relative Gaussian output noise. At 2% noise, training size ranges over $64, 96, 128, 256, 512$. Maximum fitted degree is varied around the true degree, and a rank-three cubic target is fitted with Waring ranks $1, 2, 3, 5$. Direct blocks are whitened using the training design. A validation-selected group-lasso path acts across realized degree blocks, and a degree is called active when its reference norm is at least 5% of the largest block. The independent ray fit uses 64 directions and 21 signed nodes. Table 1 separates highest-degree recovery from exact support.

Table 1: Synthetic recovery and identified failure modes. "Exact" means the entire active-degree support is correct.

Setting	Trials	Highest degree	Exact	Boundary exposed
Noise free	21	21/21	21/21	Maximum test error 6.8×10^{-16}
1–10% output noise	27	27/27	25/27	Two small spurious degree-two blocks at 10%
64 samples, 2% noise	6	3/6	2/6	All six designs flagged rank deficient
128 samples, 2% noise	6	6/6	6/6	Experiment-specific recovery threshold
Signed rays, clean	12	12/12	12/12	Independent feature construction
Signed rays, 1% noise	12	12/12	12/12	Independent noisy queries

Burgers and heat measurement. The input field is

$$a_c(x) = c_1 \sin x + c_2 \cos x + c_3 \sin 2x + c_4 \cos 2x, \tag{37}$$

with coefficient directions uniform on $\mathbb{S}^3 \subset \mathbb{R}^4$ and radii uniform in $[0.2A, A]$. The output contains the real sine/cosine coefficients of the first six positive Fourier modes. The $3 \times 3 \times 3$ physical grid is

$$A \in \{0.1, 0.4, 0.8\}, \quad \nu_{\text{visc}} \in \{0.10, 0.05, 0.02\}, \quad T \in \{0.05, 0.20, 0.80\}. \tag{38}$$

Three seeds and resolutions $256, 512, 1024$ give 243 Burgers fits. The primary statistic uses the 81 fits at resolution 512. Each fit uses 384 training, 128 validation, 128 test, and 256 reference inputs.

The direct estimator retains degrees through five. The ray estimator uses 48 new directions and 17 Chebyshev nodes $t_i \in [-1, 1]$, producing physical ray points $At_i v$ with scalar amplitudes in $[-A, A]$. Thus one direct profile replicate uses 768 operator evaluations for fitting, penalty selection, and reference norms, plus 128 held-out evaluations for fit error; one ray replicate uses $48 \times 17 = 816$ independent evaluations. The prospective score averages three such replicates. It is outcome-blind with respect to residual training, but it is neither label-free nor cheaper than using already known physical parameters.

We solve the zero-mean viscous equation through the Cole–Hopf transform. A positive periodic heat-kernel quadrature evaluates the transformed field and its analytic derivative; Fourier analysis supplies the reported output modes. At the strongest anchor, the first 32 output modes at resolutions 512 and 1024 have relative discrepancy at most 4.49×10^{-15}. The matched heat equation uses the identical input distribution, viscosity, time, grids, and output coordinates. Every one of its 81 primary-resolution fits selects only degree one.

The Burgers study predates the energy estimand and uses

$$p_{k,1}(\tau) = \frac{\tau^k M_k^{1/2}}{\sum_{j=1}^{K} \tau^j M_j^{1/2}}, \qquad \kappa_1(\tau) = \sum_k k p_{k,1}(\tau). \tag{39}$$

Its values are not pooled numerically with $p_{k,2}$. For a fixed surrogate, the same differentiation as in Appendix A.1 gives $\mathrm{d}\kappa_1/\mathrm{d}\log\tau = \mathrm{Var}_{p_{\cdot,1}}(k)$. The empirical surrogate is refitted at each amplitude, so its observed monotonicity is compatible with, but not implied by, this transport identity.

B.2 Approximation oracles and statistical sensitivity

For every Burgers condition, five nested dense polynomial maps retain degrees one through K', with $K' = 1, \ldots, 5$. With four inputs, their cumulative feature counts are $5, 15, 35, 70, 126$ after including the constant nuisance term; the number of fitted output coefficients therefore grows with K'. This is an approximation oracle, not a capacity-matched architecture comparison.

For evaluation domain d, the two prespecified marginal values are absolute error reductions

$$G_{1:5}^{(d)} = e_1^{(d)} - e_5^{(d)}, \qquad G_{3:5}^{(d)} = e_3^{(d)} - e_5^{(d)}. \tag{40}$$

Each of 81 regime–seed records uses 384 training coefficients and 256 fresh coefficients in each of ID, near-OOD, and far-OOD shells. A complete rerun uses new coefficient draws. The eight prespecified ID/near correlations pair direct or ray κ_1 with $G_{1:5}$ and direct or ray $p_{\geq 3,1}$ with $G_{3:5}$; they range from 0.986 to 0.998 in the rerun. Growing capacity and the missed 15% strong-anchor threshold limit the interpretation to approximation value, not module efficiency.

The original study met three of five criteria. Its far-OOD correlations were only 0.635–0.758, and at the strongest anchor the median relative error increased from 0.5463 at degree one to 0.6151 at degree five. The new-draw replication was deliberately restricted to ID and near OOD; it neither reran nor superseded that far-shell failure.

Physical-regime statistical unit. The reported correlations above treat the 81 regime–draw records as the units of the prespecified matrix statistic. To test whether repeated coefficient draws create a misleadingly large effective sample, we also average the three draws within each $(A, \nu_{\mathrm{visc}}, T)$ cell and recompute every association on the resulting 27 physical regimes. We then draw 100,000 nonparametric bootstrap samples of 27 whole regimes and delete each regime once. Table 2 shows that aggregation does not explain the association.

All 800,000 bootstrap draws are defined. An interval of $[1, 1]$ reflects perfect rank monotonicity in the observed regime means; resampling cannot introduce unseen inversions. These intervals describe stability on the fixed $3 \times 3 \times 3$ grid used to define the associations, not design-based uncertainty or prospective transfer to new physical regimes.

Sensitivity to energy weighting. The primary predictive analyses were fixed under the RMS-normalized member $q = 1$. To assess dependence on that choice without creating a new outcome study, we recomputed $p_{k,2} = M_k / \sum_{j \geq 1} M_j$ by squaring the stored block RMS norms. This post-hoc analysis makes no new PDE queries and leaves every trained-model outcome unchanged.

Table 2: Regime-level sensitivity of the replicated $q = 1$ associations. Intervals resample complete physical regimes; LOO is the range over all 27 single-regime deletions.

Domain	Pairing	ρ	Bootstrap 95%	LOO range
ID	direct κ_1 vs. $G_{1:5}$	1.0000	$[1.0000, 1.0000]$	$[1.0000, 1.0000]$
ID	ray κ_1 vs. $G_{1:5}$	1.0000	$[1.0000, 1.0000]$	$[1.0000, 1.0000]$
ID	direct $p_{\geq 3,1}$ vs. $G_{3:5}$	0.9994	$[0.9946, 1.0000]$	$[0.9993, 0.9998]$
ID	ray $p_{\geq 3,1}$ vs. $G_{3:5}$	1.0000	$[1.0000, 1.0000]$	$[1.0000, 1.0000]$
Near	direct κ_1 vs. $G_{1:5}$	0.9933	$[0.9574, 1.0000]$	$[0.9925, 0.9993]$
Near	ray κ_1 vs. $G_{1:5}$	0.9933	$[0.9574, 1.0000]$	$[0.9925, 0.9993]$
Near	direct $p_{\geq 3,1}$ vs. $G_{3:5}$	0.9957	$[0.9714, 1.0000]$	$[0.9952, 0.9993]$
Near	ray $p_{\geq 3,1}$ vs. $G_{3:5}$	0.9963	$[0.9741, 1.0000]$	$[0.9959, 1.0000]$

On the new-draw 27-regime matrix, the four direct/ray, ID/near-OOD correlations between $p_{\geq 2,2}$ and the squared-error-aligned gain $e_1^2 - e_5^2$ range from 0.98604 to 0.99715. The stricter $p_{\geq 4,2}$ comparison with $e_3^2 - e_5^2$ gives 0.92635 and 0.92652 for the direct estimator, versus 0.99621 and 0.99625 for signed rays. The original far-shell counterparts remain weaker, beyond the scope of the narrower replication.

For the balanced allocation study, substituting $p_{\geq 3,2}$ into the original within-PDE direct/ray average-rank rule reconstructs the same 24 conditions and the same twelve-condition selected subset. It therefore inherits, rather than independently re-estimates, the near ratio 0.98841 $[0.98301, 0.99368]$, the far ratio 1.00459 $[0.99953, 1.00983]$, and the same four omnibus components outside the joint criterion. This exact stability is a post-hoc fixed-data robustness result rather than a prospective $q = 2$ confirmation. Redesigning the candidate matrix with $q = 2$ effective degree or $p_{\geq 4,2}$ changes the selected conditions, for which no allocation outcomes exist, so their performance remains unidentified.

B.3 Prospective allocation and full-field controls

The second nonlinear family is periodic Allen–Cahn,

$$\partial_t u = \nu_{\text{visc}} \partial_{xx} u + (u - u^3), \tag{41}$$

solved by Strang splitting between exact Fourier diffusion and the exact scalar reaction flow. Separate 64-condition Burgers and Allen–Cahn candidate pools are disjoint from all labeled development matrices. Direct and signed-ray profiles are measured with three outcome-independent replicates. Within each PDE, twelve fixed quantile centers of their average rank form six low- and six high-interaction conditions. The score is the average within-PDE rank of direct and signed-ray RMS cubic-plus mass $p_{\geq 3,1}$. The physical parameters and scores fixed before outcome evaluation are included in the machine-readable matrix supplied with the paper.

The residual comparison uses ten paired data/model replicates for each of 24 conditions. Each selector activates six conditions per PDE. Near-OOD radii are 1.05–1.25 times the training boundary; the far shell is 1.30–1.50. A condition is positive when its ten-replicate mean near-OOD log ratio favors augmentation. The interaction selector is fixed before model outcomes. Physical-only and outcome-trained ridge selectors use 81 closed labeled development conditions with whole-matrix holdouts; the validation selector uses only ID validation gain. The random reference enumerates 100,000 stratified equal-count subsets.

Intervals are percentile intervals from 100,000 resamples of the ten complete replicate blocks, preserving all conditions and domains within a block. They are conditional on the fixed 24-condition matrix and do not include candidate- pool selection uncertainty. Ratios are obtained by exponentiating mean paired log ratios. The original decision required all 16 criteria in Table 3; four fail, so the joint criterion is not met.

Physical-only and outcome-trained ranks select exactly the same twelve conditions as interaction order. This equality is shown rather than counted as an incremental advantage. The ID-validation selector selects only 5/12 positive conditions and has near ratio 1.02733 with interval $[1.01062, 1.04574]$. The far-shell reversal is reported adjacent to the near-shell result and is not repaired by a post-outcome cutoff.

Table 3: Complete prospective allocation decision. Favorable near-OOD components do not override the unmet joint criterion.

Prespecified criterion	Outcome
Balanced prevalence; at least 9/12 selected positives; enrichment ≥ 0.20	PASS: $14/24$; $10/12$; $+0.250$
Pooled AUROC ≥ 0.80; each-PDE AUROC ≥ 0.75	FAIL: 0.7821 pooled; 0.700 Allen–Cahn
Selected near ratio ≤ 0.99 and interval below one	PASS: 0.98841; $[0.98301, 0.99368]$
Selected far ratio ≤ 0.995 and interval below one	FAIL: 1.00459; $[0.99953, 1.00983]$
Near/far paired-positive fractions	PASS: $0.8583/0.7417$
Random support and near-value tails ≤ 0.05	PASS: $0.00706/0.00007$
Noninferiority to learned and validation selectors; exact fallback	PASS

Table 4: Full-field evidence chain. Earlier favorable means do not establish the final structural claim.

Stage	Observed result	Consequence
Two Allen–Cahn studies	Composed/FNO ratios 0.9729 and 0.9831; both 95% intervals cross one	Individual superiority not established; pooled 0.9801 is secondary
Topology and cross-PDE controls	Degree-≤ 3 matches or outperforms high order on an already-open split; high-order/FNO ratio 1.0094	High-order necessity not supported by corrective development
Residual calibration	The full-field residual-norm criterion is not met; the earlier residual is pointwise; post-stop exploratory ratios $0.838/0.996$ are descriptive rather than inferential	Confirmatory training was not started; optimization remains plausible
Saturation-first comparison	The nonlocal construction is verified; none of three fresh FNOs reaches saturation	Performance arms were not trained; the comparison remains open

Full-field evidence and the saturation-first test. Table 4 separates the scientific meaning of four progressively stronger full-field controls.

For a conditional FNO F_ϕ and the residual above, the intended comparison was

$$\widehat{\mathcal{G}}_\theta(u) = F_\phi(u, \theta) + \alpha R_\psi(u, \theta), \qquad 0 \leq \alpha \leq 1, \tag{42}$$

with one globally validation-selected α. Pointwise, degree-≤ 3, generic-nonlinearity, and optimizer-time-matched continued-FNO controls were planned. The residual uses two $\{1, 3\}$ stages, 64 heads per stage, eight real-even spectral coordinates, and 2,224 trainable scalars on periodic 32×32 Swift–Hohenberg fields.

The comparison required three fresh FNOs to saturate: after 6,000–12,000 AdamW updates, $(b_{t-8} - b_t)/b_{t-8} < 0.005$ over eight validation intervals, where b_t is the validation error used by the stopping test at checkpoint t. All three instead attain their lowest validation error at 12,000 updates, with errors 3.4211×10^{-5}, 3.0701×10^{-5}, and 4.1421×10^{-5} and rolling improvements 0.15383, 0.15762, and 0.17736. Thus $0/3$ pass, and training for the planned residual, control, and continued-FNO comparisons did not proceed. Algebraic analysis and implementation checks establish the construction's feasibility. Its predictive accuracy, including the value of nonlocal access, degrees five through nine, and homogeneous sectors, remains untested in this comparison.

C Evaluation, reproducibility, and solution visualizations

For prediction $\widehat{y}_i$ and target y_i, every reported relative error uses

$$e = \frac{\left(\sum_i \|\widehat{y}_i - y_i\|_2^2\right)^{1/2}}{\left(\sum_i \|y_i\|_2^2\right)^{1/2}}. \tag{43}$$

For paired methods A and B, geometric ratios are

$$\exp\left(\frac{1}{S}\sum_{s=1}^{S}\log\frac{e_{A,s}}{e_{B,s}}\right). \tag{44}$$

The sampling unit is the complete independent replicate, not a condition evaluated by the same fit. Spearman coefficients use the full matrix and standard mid-ranks for ties. The anonymous release contains the reported results, analysis code, and main-figure source data. Figure 1(a) is analytic; the other main-figure values come from the fixed result set. The reference fields below are separately generated numerical illustrations. In the saturation-first comparison, the baseline models did not meet the saturation criterion, so training and accuracy evaluation of the proposed residual models did not proceed.

Reference solution morphology. Figures 3 and 4 show reference solutions to make the spatial dynamics underlying the operator tasks concrete. These illustrations use the existing numerical solvers and fixed initial data; they are not neural-operator predictions or additional tests of allocation performance.

For Burgers and heat, we use the same coefficient direction $(0.6, -0.3, 0.4, 0.6)/\sqrt{0.97}$ and viscosity $\nu = 0.02$. Normalizing the field by A separates changes in shape from simple amplitude scaling. The heat solution retains its smooth profile, whereas stronger Burgers advection displaces extrema and steepens the descending front. Cole–Hopf solutions use 512 spatial points and 81 displayed times. Comparison with 1024 points at $t = 0.01, 0.2, 0.8$ gives relative ℓ_2 discrepancies below 10^{-12} for these fields.

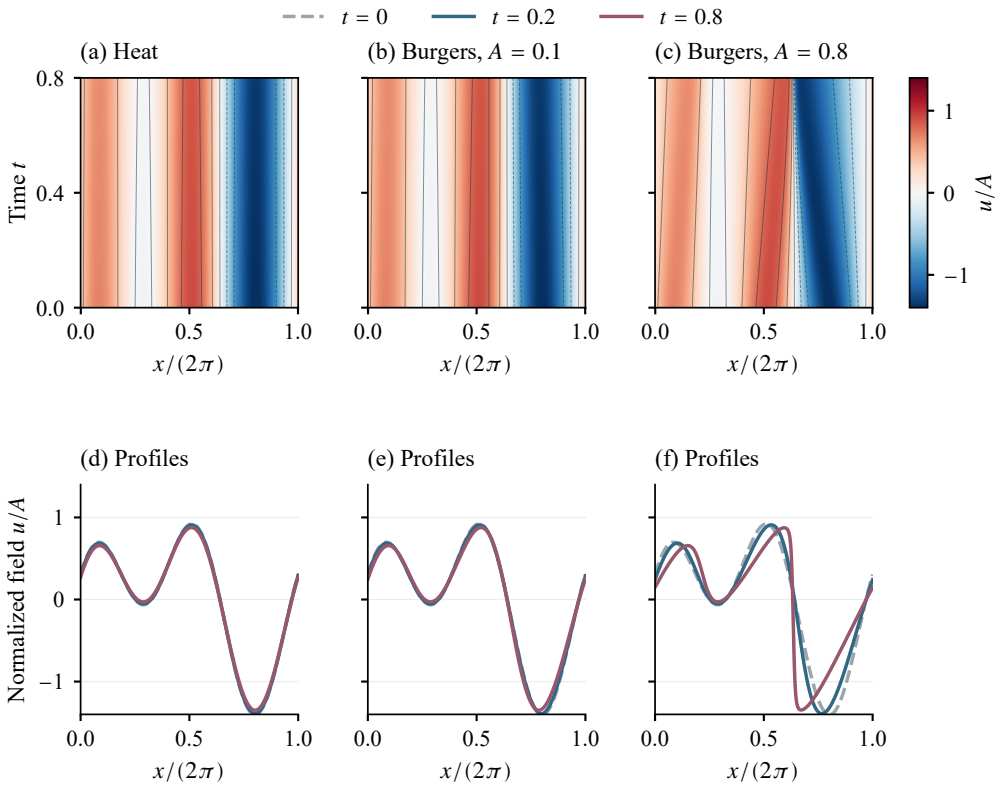


Figure 3: **Reference solution evolution on the periodic interval.** (a–c) Normalized space–time fields at fixed viscosity $\nu = 0.02$. The heat solution is amplitude independent after normalization; Burgers solutions use $A = 0.1$ and 0.8. (d–f) Matched spatial profiles at $t = 0, 0.2, 0.8$. All panels share the same initial coefficient direction.

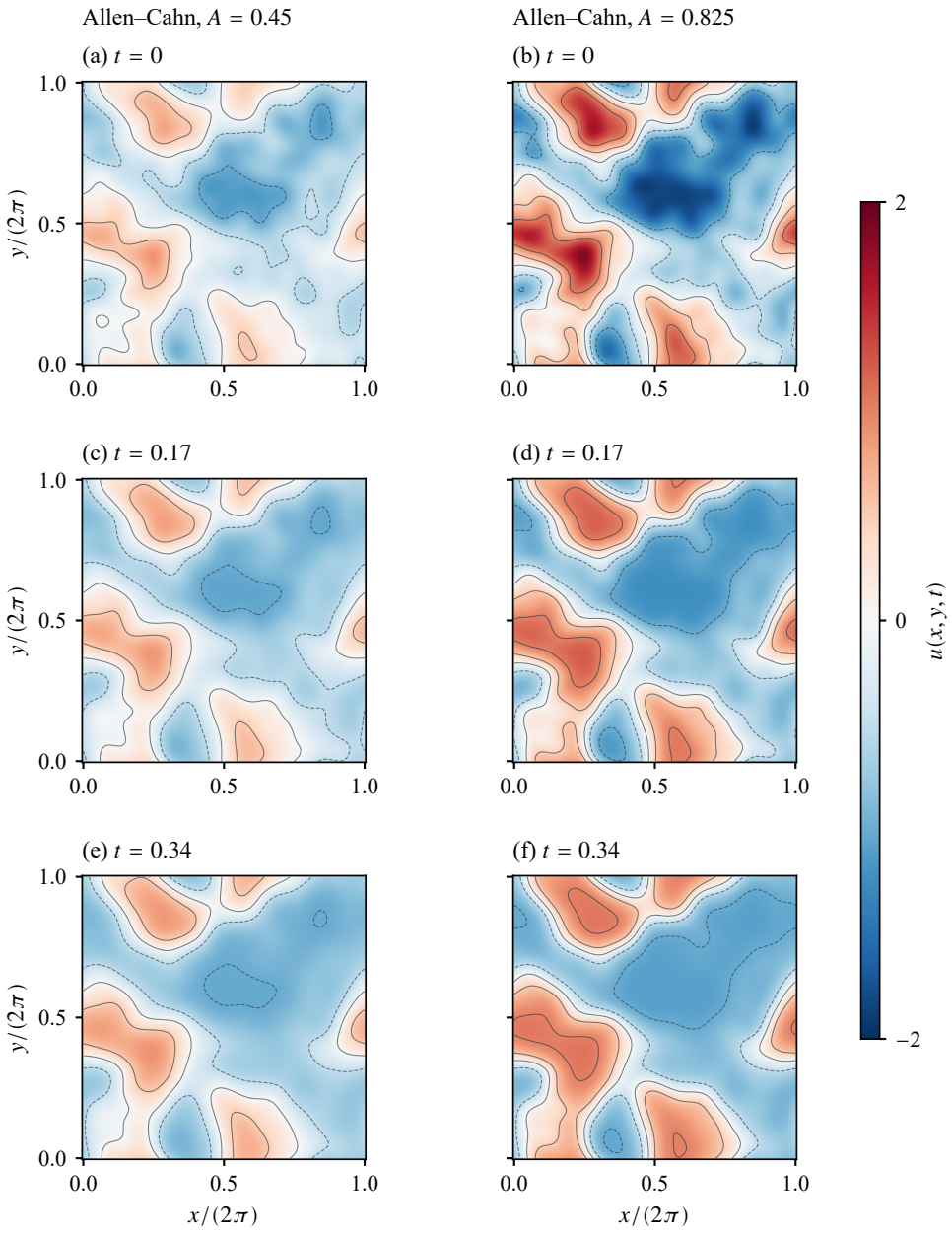


Figure 4: **Two-dimensional Allen–Cahn reference fields.** Columns use amplitudes $A = 0.45$ and 0.825 with a common initial realization. Rows show $t = 0, 0.17, 0.34$ at $\nu = 0.055$ and $\lambda = 2$. Color gives the solution value on a common scale; contours at $u = -1, -0.5, 0, 0.5, 1$ expose the evolving geometry.

For the two-dimensional Allen–Cahn equation $u_t = \nu\Delta u + \lambda(u - u^3)$ on $[0, 2\pi)^2$, we fix $\nu = 0.055$, $\lambda = 2$, and amplitudes $A = 0.45, 0.825$ from the full-field study. A common random-field realization isolates the amplitude effect. The initial field follows the study's Fourier-filtered construction with cutoff ten and seed 20260925. We spectrally resample this field from 32×32 to 128×128 and apply the existing Strang solver with maximum step 0.0025. Simultaneously doubling the spatial resolution and halving the time step changes the final fields by less than 1.2×10^{-5} in relative ℓ_2 norm. A shared color scale shows both smoothing of small spatial features and the evolution of broad positive and negative regions.