# EMERGENCE OF DYNAMIC SIMILARITY IN NEURAL REPRESENTATIONS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Physical systems can look very different in dimensional coordinates while still obeying the same governing dynamics. When neural networks learn such systems, do their internal representations organize around the dimensionless relationships identified by classical physics? We study neural networks trained on publicly available physics datasets from seven dynamical systems spanning transport, mechanics, reaction–diffusion, and waves. We find that training consistently reorganizes hidden representations around the dimensionless physical relationships that govern the observed dynamics. This organization can strengthen in two ways: by emphasizing changes that alter the underlying physics or by becoming less sensitive to equivalent dimensional descriptions. Yet independently trained networks need not share the same neural coordinates, suggesting that representational convergence can occur at the level of physical relationships rather than a common latent basis. These results motivate the broader hypothesis that independently learned representations may converge on the relationships that structure the underlying physical world.

## 1 INTRODUCTION

> *"It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience."*
>
> Albert Einstein (1934).

The Platonic Representation Hypothesis demonstrates an interesting phenomenon: as vision and language models scale, the similarities they assign to observations become increasingly alike (Huh et al., 2024). This raises a question for scientific machine learning. When independently trained networks learn physical systems, do they develop a shared way of representing the underlying physics? And if such structure appears, does it resemble the relationships already used in classical theory?

Physical observations are partial views of a much larger space of possible states and conditions. Their variation is constrained by laws, symmetries, and relations among quantities. This makes physics a concrete setting for asking what independently learned representations might share. A reusable relation can explain many observations even when each network encodes it in different coordinates. Representation learning may favor such common structure, although neither simplicity nor task performance guarantees that it will. Classical physics gives this broad question a testable reference: dimensional analysis identifies combinations that change normalized behavior and transformations that preserve it. These known relationships let us study part of the organization behind observed trajectories while distinguishing physical agreement, shared coordinates and accurate prediction.

Systems with different sizes, time scales or material properties can exhibit the same normalized behavior. *Nondimensionalization* expresses this relationship through combinations whose units cancel (Buckingham, 1914). Distance traveled divided by system length, for example, describes physical progress independently of the units used. These combinations distinguish changes in dynamics from changes in dimensional description.

Representation learning can expose latent variables that organize observed behavior, including in recurrent networks trained on next-observation prediction (Recanatesi et al., 2021). *Hypothesis. Training neural networks on physical dynamics makes hidden representations increasingly selective*



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.

*for the dimensionless dependency spaces governing those dynamics, even though those spaces are neither supplied as representation targets nor imposed on hidden layers.* We distinguish physical dependence, shared neural coordinates and accurate physical behavior: evidence for one does not establish the others.

A physical quantity may be decodable because a layer preserves its inputs, while distributed information may evade a simple readout. We therefore distinguish retained information from learned organization by testing both readouts and responses to physical transformations (Hewitt & Liang, 2019).

Dimensional analysis defines the physical reference. We examine how training changes hidden dependence, how finite descriptions of equivalent physics relate, and what independently trained models share despite different coordinates. Physical evolution at the output then tests how accurately each network uses this organization.

## 2 From observations to physical variables

### 2.1 Representations

A *representation* consists of the activations at a specified layer, which the remaining network uses to solve its task (Bengio et al., 2013). Writing $\widehat{y} = g_\ell(z_\ell(X))$, $z_\ell$ may be a vector, token sequence, or spatial feature field. The input $X$ specifies an initial state, dimensional parameters, and a requested time, with spatial coordinates where needed. We call the complete activations in the model's own coordinates its *native features*, before auxiliary projection or alignment. Comparisons depend on which coordinate transformations their measurements disregard (Kornblith et al., 2019).

### 2.2 Probes

A *probe* is an auxiliary estimator fitted to frozen activations; the physical predictor remains unchanged (Alain & Bengio, 2016). Linear probes test whether physical quantities are accessible to a restricted readout. Initialization and input controls distinguish ingredients already available from organization developed through training (Hewitt & Liang, 2019). Native sensitivity tests complement these readouts by measuring which coordinated physical changes move the features themselves.

### 2.3 Dimensional analysis and physical similarity

A complete, dimensionally homogeneous physical relation can be expressed through dimensionless products (Buckingham, 1914). For $n$ included quantities whose dimensional matrix $D$ has rank $r$, its null space has dimension $n - r$. Each vector $a_j$ in that null space defines

$$\Pi_j = \prod_{i=1}^{n} q_i^{a_{ij}}, \qquad Da_j = 0. \tag{1}$$

The columns of $D$ give each quantity's powers of base units; $a_{ij}$ is the exponent of $q_i$ in product $j$. The zero condition cancels every unit. Dimensional analysis therefore identifies equivalence classes of dimensional descriptions: distinct inputs can describe the same normalized physical problem when the complete controls match. The theorem restricts allowable dependencies but leaves the function relating these products unknown. Governing equations, observations, or learning must supply that function.

For a translating profile, speed $c$, time $t$ and domain length $L$ give progress $ct/L$. Matching progress and normalized initial shape identifies the same normalized transport problem. For Burgers, the full dimensional set includes velocity scale $U$, length $L$, kinematic viscosity $\nu$ and time $t$. Its two combinations $\mathrm{Re} = UL/\nu$ and $\tau = tU/L$ describe regime and progress. The principal local assay holds $U$ fixed; this restriction does not remove $U$ from the underlying dimensional analysis. Equivalent descriptions must also preserve normalized initial conditions, geometry, boundaries and the complete relevant controls. One matching named number alone is insufficient.

Dimensionless coordinates are nonunique (Bakarji et al., 2022). We compare their dependency spans; individual neurons need not equal named quantities.

For positive quantities and fixed references,

$$\log(\Pi_j/\Pi_{j,*}) = a_j^\top \log(q/q_*). \tag{2}$$

Thus exponent vectors are the gradients of log groups, providing the dependency directions used in Section 3. A supported assay chart may vary fewer controls than the full parameterization; $n - r$ counts neither tested controls nor hidden dimensions. Signed mechanical coordinates retain their actual local Jacobian, without logging signed momentum.

Known similarity actions structure our dimensional examples, but physical groups are not predictor channels, losses or hidden targets. Dimensional consistency is not imposed on ordinary hidden layers. These physical combinations identify coordinated input changes that preserve the normalized problem.

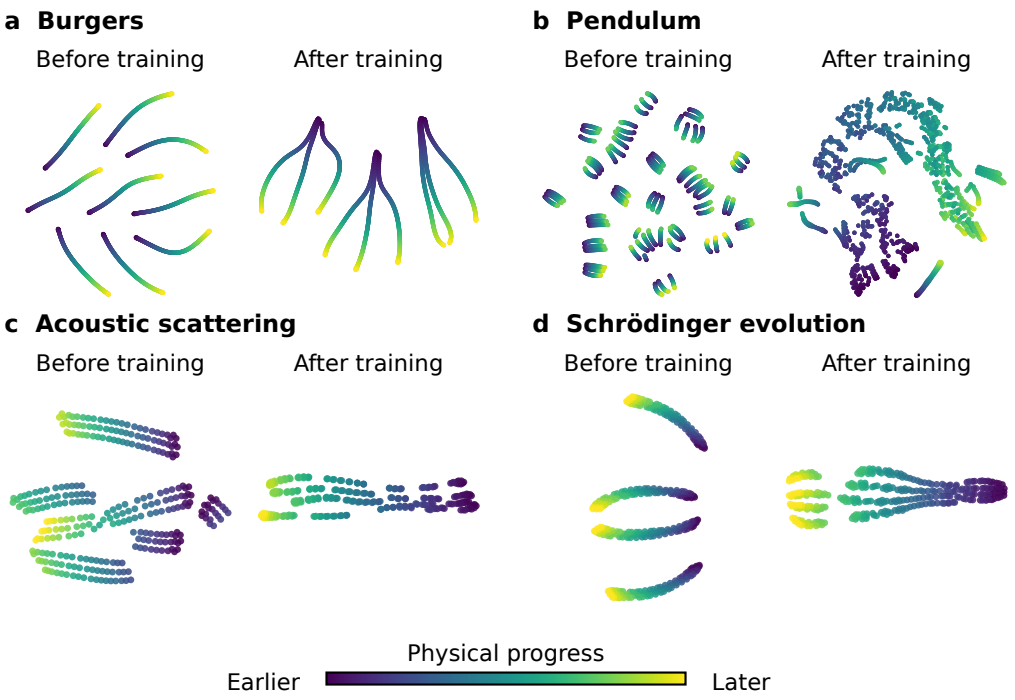



**Figure 1: Representative views of learned physical organization.** Hidden responses before and after training are colored by physical progress for four representative systems. The views illustrate how representation structure changes through training; independently fitted projections are qualitative and their axes and widths are not directly comparable. Figure 2 provides the native-feature analysis across all seven systems, with the complete montage in Appendix F.2.

We next measure whether training makes hidden responses reflect the distinction between physical changes and equivalent descriptions.

## 3 Learning the Governing Relationships

Fourier neural operators (FNOs) and residual CNNs predict fields (Li et al., 2021; Kovachki et al., 2023); residual MLPs and scalar-token Transformers predict mechanical states. We compare complete middle-layer features at initialization and checkpoints selected by ordinary validation prediction error. The synthesis contains 38 trained fits from 32 distinct random initializations; six Burgers pairs share initial weights across their two training arms. Systems, physical parents, timestamps and transformed descriptions are not draws from one common population. Table A2 gives seeds, parent roles, hooks and assay populations. All descendants of an initial condition retain its split.

| System | Predicted state | Physical relationships probed |
|---|---|---|
| Advection | Transported profile | Progress $ct/L$ |
| Burgers | Velocity field | Regime $\mathrm{Re} = UL/\nu$; progress $tU/L$ |
| Pendulum | Angle, momentum | Initial state $\theta_0, j_0$; natural time $t\sqrt{g/l}$ |
| Gray–Scott | Concentrations $A, B$ | Progress $\kappa t$; feed/removal contexts |
| Compressible flow | Coupled fields | $\mathrm{Re}$, $\mathrm{Ma}$; progress $tU/L$ |
| Acoustic scattering | Pressure, velocity | Acoustic progress $c_{\mathrm{ref}}t/L_{\mathrm{ref}}$ |
| Schrödinger | Complex wavefunction | Progress $\omega t$; fixed $\hbar/(m\omega L^2)$ |

**Table 1: Physical systems and relationships examined in the representation assays.** Burgers, pendulum, and compressible flow expose non-temporal physical variation in addition to progress; the local Gray–Scott, acoustic, and Schrödinger assays primarily probe progress. Exact dimensional variables and similarity transformations are given in Appendix A.

### 3.1 Measuring physical dependence

The native Jacobian describes how complete hidden features respond to proportional dimensional-input changes. Write $s_i = \log(q_i/q_{i,*})$, and collect the exponent directions $a_j$ from Section 2 in $B$. If features depend on these groups at fixed normalized initial context, then

$$\xi = B^\top s, \qquad z = h(\xi), \qquad J_z = J_h B^\top. \tag{3}$$

Here $J_z = \partial z/\partial s$ differentiates the complete input construction. Each assay restricts these directions to its registered supported chart and uses its original orthonormal coordinate metric. Signed mechanical inputs follow the actual local Jacobian. Within that chart, the physical projector $P$ selects changes that alter governing controls; its complement selects changes preserving them. We measure their respective response energies and physical fraction:

$$E_{\mathrm{phys}} = \|J_z P\|_F^2\,, \quad E_{\mathrm{scale}} = \|J_z(I-P)\|_F^2\,, \quad S_{\mathrm{phys}} = \frac{E_{\mathrm{phys}}}{E_{\mathrm{phys}} + E_{\mathrm{scale}}}. \tag{4}$$

The physical share describes sensitivity in the network's native feature geometry, unchanged by orthogonal feature transformations and uniform scaling. It can increase through larger physical response, smaller description response, or both. We test specificity against dimension-matched input subspaces and robustness under bounded nonuniform feature transformations (Appendix C.2). Principal angles and singular values separately test coverage of the physical span (Constantine et al., 2017); they remain sensitive to feature scaling. Derivatives include the complete input construction. We preserve each estimator: a mean of pointwise fractions differs from a fraction of pooled energies.

### 3.2 Evidence across experiments

Figure 2 compares complete native responses across physical systems. Its columns retain each recorded population and pointwise or pooled estimator; projections do not enter these measurements.

Training increases alignment with the physical dependency structure supported by the available data in all seven physical systems (Figure 2a). The advection, Gray–Scott, acoustic and Schrödinger local assays primarily test normalized progress. Burgers, pendulum mechanics and compressible flow also expose non-temporal physical directions.

Physical organization extends beyond elapsed progress in these richer charts (Figure 2c). Clock-removed Burgers shares are $0.854$–$0.942$, against expectation $0.5$. Pendulum MLPs retain especially strong non-temporal dependence, while both compressible architectures retain more than $0.97$ of their residual sensitivity in the non-clock physical span. Each uses its remaining physical rank and reference expectation; these fractions do not define a pooled score or common neural coordinates.

Specificity varies across physical systems (Figure 2b). We compare governing spans with 2,048 fixed random orientations of equal dimension, conditional on the measured derivative. Reference percentiles are not population confidence levels. Clock-removed Burgers dependence is not uniformly beyond the central $95\%$ reference range. Acoustic and Schrödinger clock-direction shares increase, but neither system exceeds the $97.5$th percentile of the common rank-one reference. Complete comparisons are in Appendix D.1.

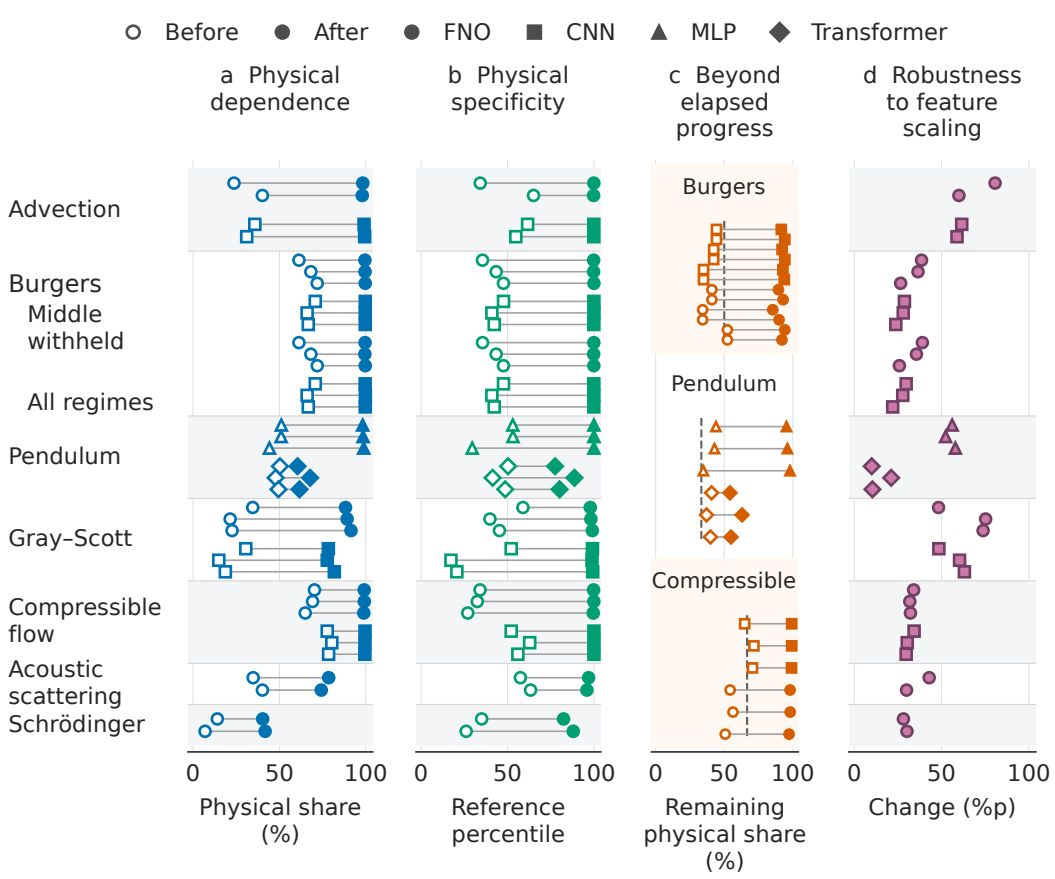



**Figure 2: Training reorganizes representations toward physical dependencies across seven systems.** (a) Physical-direction sensitivity increases after training. (b) Trained representations preferentially align with the physical subspace relative to random subspaces of the same dimension. (c) Burgers, pendulum, and compressible flow retain substantial physical dependence after elapsed progress is removed. (d) The training-induced increase persists after shared feature normalization. Panels a, b, and d include all 38 fits; panel c includes the 24 fits whose physical span contains additional non-temporal directions. Advection and compressible flow use pooled energy ratios; other systems average pointwise shares. Panel d uses matched subsets for advection, Burgers, and compressible flow. Burgers training arms share initial weights; Appendix B gives model ancestry and evaluation populations.

All 38 fits retain a positive paired increase under shared calibration-based channel normalization (Figure 2d). A sufficient bound certifies a positive increase for 29 of 38 fits under every invertible linear feature transformation with condition number at most two. The four acoustic/Schrödinger fits add no certificate at this threshold; their overlapping bounds are inconclusive. The full conditioning grid and measured gain assignments remain in Appendix D.1; sampled positive changes do not establish invariance to every invertible encoding.

Our training distributions include similarity-structured coverage, so these experiments do not establish that the same organization would arise under arbitrary dimensional sampling. We test whether ordinary hidden layers internalize this physical organization without dimensionless representation targets or constraints.

### 3.3 Finite descriptions of the same physics

Local sensitivities identify the directions a representation emphasizes. Finite rescalings show how that organization appears across descriptions of the same normalized problem. Figure 3 pairs a fixed

pendulum example with complete-native Burgers comparisons of physical separation and description effects.

The full-native comparison shows two routes to stronger relative organization. FNO description spread grows roughly 5–16 times, while temporal and regime separation grow faster. CNN description spread shrinks to 0.11–0.49 of initialization; some regime separations also shrink, but less than their corresponding description spread. All registered comparisons retain this relative ordering. Thus stronger physical organization need not mean absolute invariance.

Small equivalent descriptions also develop more closely aligned directions of neural response as supplied physical progress changes. This consistency is range dependent: wider mechanical alignment worsens (Appendix C.5). These observations complement input-Jacobian sensitivity without defining an autonomous neural flow.

Acoustic small-description relative distances fall from 0.21–0.29 to 0.030–0.037, while absolute distances grow. Schrödinger output residuals remain nonzero after removing the known amplitude factor. Compressible flow gains relative physical energy while its leading physical subspace remains poorly aligned. These observations distinguish relative emphasis, absolute contraction, coverage of physical directions and output covariance (Appendices C and D).

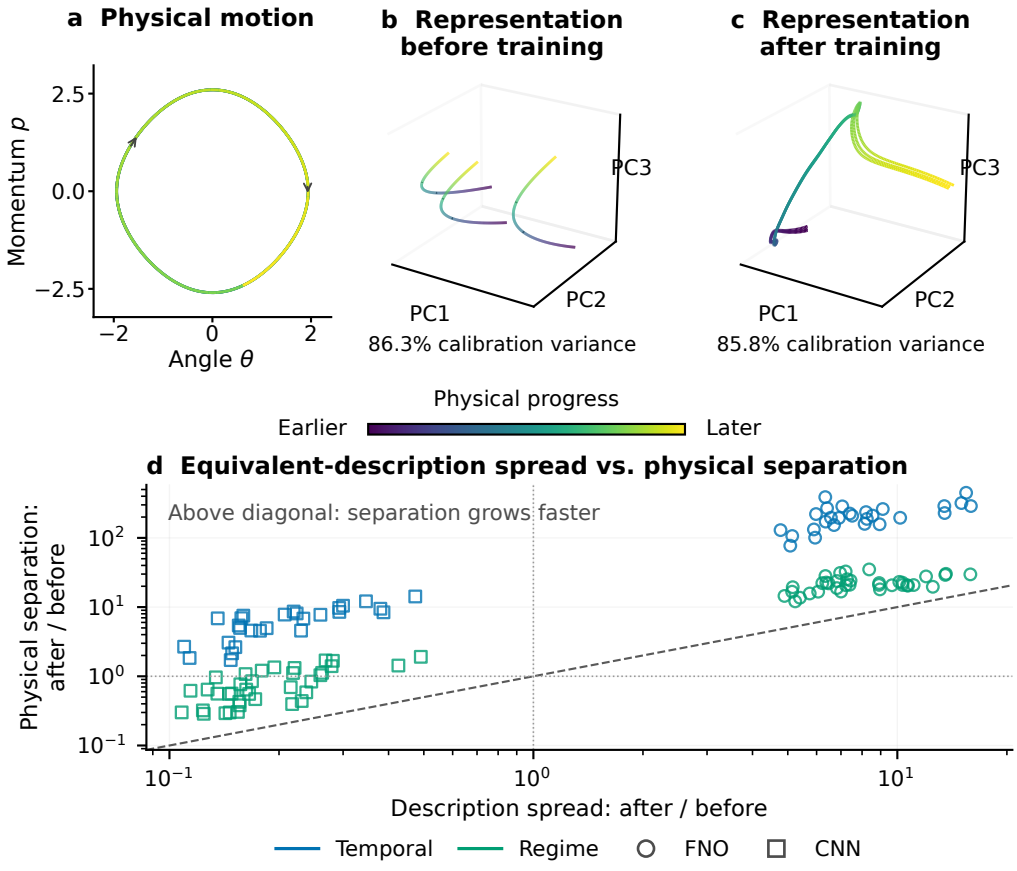



**Figure 3: Physical organization can strengthen through separation or contraction.** (a–c) A pendulum trajectory and its representation under three physically equivalent dimensional descriptions before and after training. (d) Burgers compares the training-induced change in equivalent-description spread with the change in temporal or regime separation across 126 native-feature comparisons. Above the diagonal, physical separation grows faster than description dependence; left of one, equivalent descriptions contract. FNOs primarily follow the first route, while CNNs often follow the second. Separate-stage PCA views in b–c are illustrative and do not define a shared coordinate system.

# 4 SHARED PHYSICS WITHOUT SHARED COORDINATES

If physical convergence concerns relationships rather than a shared neural basis, independently trained models need not admit one transferable linear coordinate system. We test this by fitting label-free linear correspondences on common calibration inputs, freezing those maps, and transferring physical readouts between models. Initialization and raw-input controls distinguish shared information from organization acquired through training (Appendix C).

Burgers reference-description transfer declines after training, although normalized-time transfer improves on some rescaled domains. Positive within-model readouts can coexist with strongly negative transferred $R^2$, so recurring physical dependence need not yield a common linear coordinate system (Kornblith et al., 2019; Bansal et al., 2021). Acoustic directional clock transfer instead rises from $0.469/0.574$ to $0.911/0.932$; Schrödinger transfer also improves. Shared acoustic clock decoding is already $0.924$ at initialization. Changing parent pairing preserves directional transfer, whereas reversing time disrupts it, identifying the transferable temporal component.

Auxiliary nonlinear condition-geometry results were mixed and partly reflected input structure; complete analyses are retained in the accompanying evidence package. Common physical organization therefore need not appear as a common neural coordinate system. We next ask how accurately networks use that organization to reproduce physical behavior.

# 5 FROM PHYSICAL ORGANIZATION TO PHYSICAL BEHAVIOR

## 5.1 PHYSICAL EVOLUTION

Figure 4 shows the motion of a complex wavefunction and pressure scattered by heterogeneous material. Signed residuals locate the differences from the archived fields. Schrödinger predictions track the displayed evolution with mean complex-field error $0.193$–$0.196$ over forty times and $0.198$–$0.202$ over the twenty-nine intermediate times. Acoustic error varies more strongly between supervised observations: MSE is $0.768$–$0.807$ over forty-eight intermediate frames, versus persistence $1.350$. Both systems gain clock selectivity, while their temporal accuracy differs; this is not a controlled comparison of architectural or physical difficulty.

Snapshot error and physical response are separate measurements. For a specified reference change $\Delta y$ and its predicted change, we use

$$E_\Delta = \frac{\|\Delta\widehat{y} - \Delta y\|}{\|\Delta y\|}. \tag{5}$$

This compares error with the size of the physical effect; a zero-change prediction gives one when the reference change is resolved. The acoustic comparison across supervised, intermediate and later endpoints has error about $0.91$, versus $0.336$–$0.347$ on the distinct Schrödinger increment roster. Burgers dense within-range medians are $0.133$ for FNO and $0.454$ for CNN, despite field errors $0.020/0.032$. Sparse and dense comparisons use different intervals. Chemistry likewise combines accurate short accumulated evolution with less faithful local increments and longer queries (Appendix E).

Broader description coverage improves advection responses in the frozen eight-context selected-domain confirmation, while complementary failures persist; the 16-parent exploratory population remains separate. Endpoint-trained Burgers predicts the middle regime better than its endpoint-prediction average in all six comparisons, yet paired enrichment worsens field and viscosity-response accuracy (Appendix E).

## 5.2 PROGRESS AND RECURRING STATES

Mechanical recurrence separates elapsed progress from physical state (Figure A1). Across MLP seeds, native-range output-distance agreement with physical returns is $0.877$–$0.893$, whereas middle-feature agreement is $0.016$–$0.041$, close to the clock control's $0.023$. Middle features can distinguish times whose predicted states are similar. The test keeps unwrapped angle and canonical momentum, without phase alignment or time warping. Every evaluated sequence is direct-query prediction, not a recursive rollout.

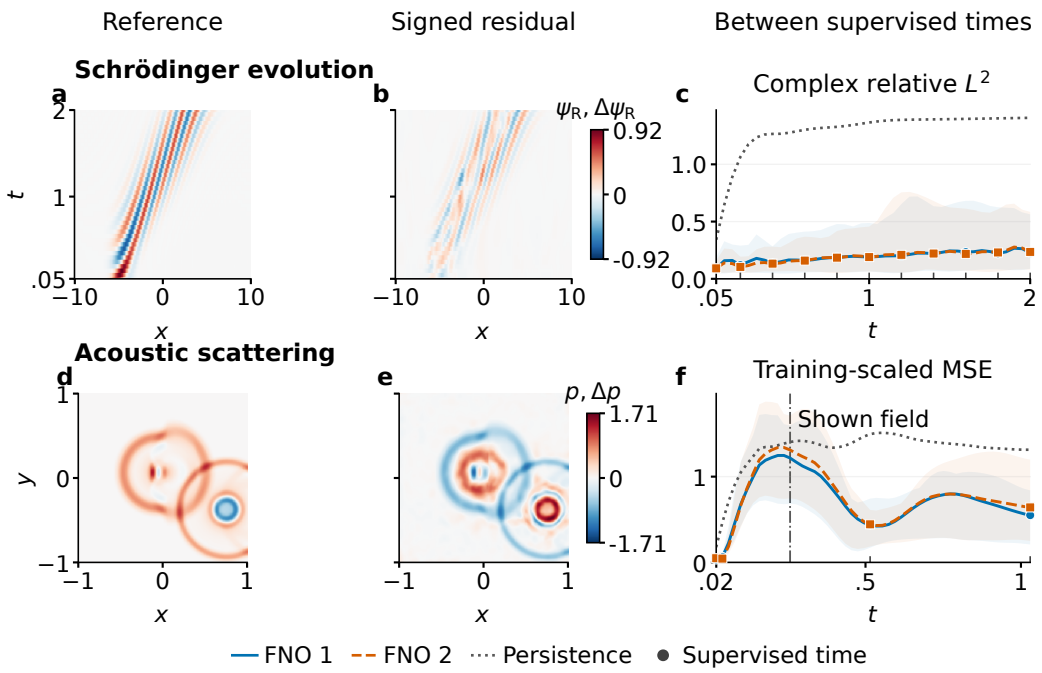



**Figure 4: From physical organization to physical evolution.** Representation-level physical organization and output fidelity are evaluated separately. Schrödinger and acoustic examples show archived fields, signed residuals, and direct-query errors between supervised times. Filled markers denote supervised times; light bands show whole-parent spread, and persistence keeps the initial state fixed. All predictions are direct queries rather than recursive rollouts.

## 6 What emerges across physical systems?

Across seven physical systems, representation learning makes native responses more selective for the physical dependency structure available in each system. The measured directions include regime and progress in Burgers, initial state and a natural clock in mechanics, and elapsed progress in the chemical, acoustic and quantum panels. Their supported physical scope remains distinct.

These results suggest that representational convergence may occur at the level of physical relationships rather than neural coordinates. Models can increasingly distinguish changes that preserve a normalized problem from changes that alter it, while using that organization differently downstream. A small set of physical controls can organize dependence throughout a high-dimensional representation without prescribing its dimension. Growing physical separation and shrinking description responses provide different routes; prediction and recurrence tests establish how their geometric expression and accuracy differ.

The dimensionless relationships are not supplied as representation targets or imposed on hidden layers, but the training distributions do expose similarity-structured physical variation. The experiments therefore establish learned organization around these relationships, while leaving open whether the same tendency would arise under arbitrary dimensional sampling.

Why this organization emerges remains unresolved. One possibility is a simplicity pressure: reusable dependence on a few physical combinations can account for many dimensional realizations through common computation. Parameter–function bias and implicit regularization support related preferences in specific settings (Valle-Pérez et al., 2019; Arora et al., 2019); task-driven abstraction and latent-variable learning offer other accounts (Johnston & Fusi, 2023; Recanatesi et al., 2021). Simplicity can also favor shortcuts (Shah et al., 2020). Our experiments do not distinguish this mechanism from alternatives. The relational interpretation connects to the motivating representation hypothesis (Huh

et al., 2024): independent models can acquire related physical dependencies without acquiring the same neural coordinates.

Dimensionless and equivariant methods impose related structure (Bakarji et al., 2022; Villar et al., 2023; Cohen & Welling, 2016); known invariants can guide test-time transformations (Lee et al., 2026). Independent fits and supported controls are limited, some initializations are shared, and most analyses are exploratory. Initialization and input comparisons keep learned dependence distinct from retained input information, while transfer and response tests assess shared coordinates and physical accuracy separately.

## 7 Conclusion

Across seven physical systems, training repeatedly reorganizes neural sensitivity toward the dimensionless physical dependency structure available in each system. In richer physical charts, substantial dependence remains beyond elapsed progress. Independent networks can acquire these related dependencies without sharing a transferable linear coordinate system, and output accuracy remains a separate test. The experiments therefore suggest convergence at the level of physical relationships rather than neural coordinates. Similarity-structured sampling leaves open whether the same organization would arise under arbitrary physical experience, and the learning mechanism remains unresolved. Whether broader physical experience drives models toward a more general relational organization of the physical world remains an open question.

## AI USE STATEMENT

OpenAI Codex was used for code implementation and debugging, literature search, and LaTeX manuscript editing.

## REPRODUCIBILITY STATEMENT

Sections 3–5 define the primary measurements and evaluation procedures. Appendix A specifies the seven physical systems, dimensional conventions, and similarity transformations; Appendix B documents data sources, splits, model configurations, checkpoint selection, and evaluation populations; and Appendix C defines the representation measurements and robustness controls. Appendices D–F report the primary per-model results, uncertainty summaries, generalization and transfer tests, and supporting analyses needed to evaluate the manuscript's claims. Appendix G records numerical validation and source-specific corrections. All experiments use publicly available simulation datasets cited in the manuscript; extended case-level results and exploratory analyses are retained in the accompanying evidence package.

## REFERENCES

ajthor. Schrödinger Equation Dataset. Hugging Face dataset, https://huggingface.co/datasets/ajthor/schrodinger-dedalus, 2025. Public harmonic-evolution archive. Pinned release recorded in the data manifest.

Guillaume Alain and Yoshua Bengio. Understanding intermediate layers using linear classifier probes. arXiv:1610.01644, 2016. URL https://arxiv.org/abs/1610.01644.

Sanjeev Arora, Nadav Cohen, Wei Hu, and Yuping Luo. Implicit regularization in deep matrix factorization. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://arxiv.org/abs/1905.13655.

Joseph Bakarji, Jared Callaham, Steven L. Brunton, and J. Nathan Kutz. Dimensionally consistent learning with Buckingham Pi. *Nature Computational Science*, 2:834–844, 2022. doi: 10.1038/s43588-022-00355-5. URL https://doi.org/10.1038/s43588-022-00355-5.

Yamini Bansal, Preetum Nakkiran, and Boaz Barak. Revisiting model stitching to compare neural representations. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/01ded4259d101feb739b06c399e9cd9c-Abstract.html.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013. doi: 10.1109/TPAMI.2013.50. URL https://arxiv.org/abs/1206.5538.

E. Buckingham. On physically similar systems; illustrations of the use of dimensional equations. *Physical Review*, 4(4):345–376, 1914. doi: 10.1103/PhysRev.4.345. URL https://doi.org/10.1103/PhysRev.4.345.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2990–2999, 2016. URL https://proceedings.mlr.press/v48/cohenc16.html.

Paul G. Constantine, Zachary del Rosario, and Gianluca Iaccarino. Data-driven dimensional analysis: Algorithms for unique and relevant dimensionless groups. arXiv:1708.04303, 2017. URL https://arxiv.org/abs/1708.04303.

DeepMind. DeepMind Hamiltonian Dynamics Suite. Public simulation dataset and source repository, 2021. URL https://github.com/google-deepmind/dm_hamiltonian_dynamics_suite. PENDULUM_COLORS archive.

Jean-Pierre Eckmann, S. Oliffson Kamphorst, and David Ruelle. Recurrence plots of dynamical systems. *Europhysics Letters*, 4(9):973–977, 1987. doi: 10.1209/0295-5075/4/9/004.

Albert Einstein. On the method of theoretical physics. *Philosophy of Science*, 1(2):163–169, 1934. doi: 10.1086/286316. URL https://doi.org/10.1086/286316.

John Hewitt and Percy Liang. Designing and interpreting probes with control tasks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 2733–2743. Association for Computational Linguistics, 2019. doi: 10.18653/v1/D19-1275. URL https://aclanthology.org/D19-1275/.

Minyoung Huh, Brian Cheung, Tongzhou Wang, and Phillip Isola. Position: The platonic representation hypothesis. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 20617–20642. PMLR, 2024. URL https://proceedings.mlr.press/v235/huh24a.html.

W. Jeffrey Johnston and Stefano Fusi. Abstract representations emerge naturally in neural networks trained to perform multiple tasks. *Nature Communications*, 14:1040, 2023. doi: 10.1038/s41467-023-36583-0. URL https://doi.org/10.1038/s41467-023-36583-0.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3519–3529. PMLR, 2019. URL https://proceedings.mlr.press/v97/kornblith19a.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://jmlr.org/papers/v24/21-1524.html.

Seokki Lee, Min-Chul Park, Giyong Hong, and Changwook Jeong. Buckingham $\pi$-invariant test-time projection for robust PDE surrogate modeling. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/d6f1710becf448dc79516dbbf33c4a79-Abstract-Conference.html.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2010.08895.

Ruben Ohana, Michael McCabe, Lucas Meyer, Rudy Morel, Fruzsina Agocs, Miguel Beneitez, Marsha Berger, Blakesly Burkhart, Stuart Dalziel, Drummond Fielding, et al. The well: A large-scale collection of diverse physics simulations for machine learning. *Advances in Neural Information Processing Systems*, 37:44989–45037, 2024.

Stefano Recanatesi, Matthew Farrell, Guillaume Lajoie, Sophie Deneve, Mattia Rigotti, and Eric Shea-Brown. Predictive learning as a network mechanism for extracting low-dimensional latent space representations. *Nature Communications*, 12:1417, 2021. doi: 10.1038/s41467-021-21696-1. URL https://doi.org/10.1038/s41467-021-21696-1.

Harshay Shah, Kaustav Tamuly, Aditi Raghunathan, Prateek Jain, and Praneeth Netrapalli. The pitfalls of simplicity bias in neural networks. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://papers.nips.cc/paper/2020/hash/6cfe0e6127fa25df2a0ef2ae1067d915-Abstract.html.

Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Dan MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. PDEBench: An extensive benchmark for scientific machine learning. In *Advances in Neural Information Processing Systems*, 2022. URL https://arxiv.org/abs/2210.07182. Datasets and Benchmarks track.

Guillermo Valle-Pérez, Chico Q. Camargo, and Ard A. Louis. Deep learning generalizes because the parameter-function map is biased towards simple functions. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1805.08522.

Soledad Villar, Weichi Yao, David W. Hogg, Ben Blum-Smith, and Bianca Dumitrascu. Dimensionless machine learning: Imposing exact units equivariance. *Journal of Machine Learning Research*, 24(109):1–32, 2023. URL https://jmlr.org/papers/v24/22-0680.html.

# A PHYSICAL SYSTEMS AND SIMILARITY TRANSFORMATIONS

Every transformation acts on the complete dimensional problem, including the initial field, coordinates, domain, coefficients, query time, and target units. The source datasets are PDEBench, the Hamiltonian Dynamics Suite, The Well and the harmonic Schrödinger release (Takamoto et al., 2022; DeepMind, 2021; Ohana et al., 2024; ajthor, 2025). The comparisons use ordinary supervised predictors of dimensional inputs. Physical groups define the subsequent reference spaces, not predictor channels or losses.

## A.1 NORMALIZED PERIODIC ADVECTION

The scalar transport equation $\partial_t u + c\partial_x u = 0$ on a domain of length $L$ has the periodic solution

$$u(x,t) = u_0(x - ct), \qquad u^*(x^*, \alpha) = u_0^*(x^* - \alpha), \qquad x^* = x/L, \quad \alpha = ct/L. \tag{6}$$

The initial profile is held fixed in relative coordinates. Here $\alpha$ is unwrapped displacement; equal displacement and equality modulo a complete circuit are separate physical comparisons. The training panel contains three speeds and nine future times before the temporal-coverage intervention.

## A.2 NORMALIZED BURGERS DYNAMICS

For $\partial_t u + u\partial_x u = \nu\partial_{xx} u$, set $u^* = u/U$, $x^* = x/L$, and $\tau = tU/L$. Substitution gives

$$\partial_\tau u^* + u^*\partial_{x^*} u^* = \mathrm{Re}^{-1}\partial_{x^*x^*} u^*, \qquad \mathrm{Re} = UL/\nu. \tag{7}$$

Thus Reynolds number and normalized time govern the normalized response at fixed initial profile and geometry. At fixed amplitude, $(\mathrm{Re}, \tau)$ and $(\tau, \nu t/L^2)$ span the same dependency space.

## A.3 COMPLETE TRANSFORMATIONS

The advection construction uses

$$x' = ax, \quad L' = aL, \quad t' = bt, \quad c' = (a/b)c, \tag{8}$$

with the scalar initial/future field unchanged at corresponding relative locations. Displacement $ct/L$ is preserved. Equal unwrapped displacement and equality after an integer number of circuits are reported as distinct tests.

For Burgers, the full similarity action is

$$x' = ax, \quad t' = (a/b)t, \quad u' = bu, \quad L' = aL, \quad U' = bU, \quad \nu' = ab\nu. \tag{9}$$

It preserves $\mathrm{Re}$ and $\tau$. The Burgers assays use the amplitude-preserving subgroup $b = 1$. The Burgers source coefficient is converted as $\nu = \nu_{\mathrm{file}}/\pi$ before constructing physical controls. Its six predictor channels are initial velocity, physical position, query time, effective viscosity, velocity scale, and length.

At $U = 1$, the normalized similarity direction is $(1, 1, 1)/\sqrt{3}$. The physical plane has orthonormal directions $(1, -1, 0)/\sqrt{2}$ and $(-1, -1, 2)/\sqrt{6}$. The finite triplet perturbation multiplies $(L, \nu, t)$ by $\exp(\pm h)$; the local directional finite difference uses a unit log-space direction. The same label $h$ therefore does not imply identical endpoints between these two procedures.

## A.4 PENDULUM MECHANICS

The mechanical study uses archived phase states from the conservative `PENDULUM_COLORS` task in the DeepMind Hamiltonian Dynamics Suite (DeepMind, 2021). We use the recorded unwrapped angle $\theta$ and canonical angular momentum $p$, together with trajectory-specific mass, length, and gravity. Images, colors, and rendering offsets are not predictor inputs. The official suite documents samples at $\Delta t = 0.05$. The archived state ordering and dimensional convention were checked against the cited source. The equations are $\dot\theta = p/(ml^2)$ and $\dot p = -mgl\sin\theta$.

The mechanical coordinates are

$$\tau_{\mathrm{m}} = t\sqrt{g/l}, \qquad j_0 = \frac{p_0}{m\sqrt{gl^3}}. \tag{10}$$

Together with the initial angle, these describe the normalized motion. At fixed $\theta_0, p_0$, the local chart $(\log m, \log l, \log g, \log t)$ supports two physical directions. Equation (10) uses signed $j_0$, not a logarithm of momentum. On a sign-preserving chart, proportional perturbations can retain the sign of $p_0$; the registered rule at exactly zero uses a dimensional momentum direction. No evaluated initial momentum is zero. The complete finite similarity action is

$$m' = \frac{b}{a^2}m, \quad l' = al, \quad g' = \frac{a}{b^2}g, \quad t' = bt, \qquad \theta' = \theta, \quad p' = p. \tag{11}$$

It preserves $\tau_{\mathrm{m}}$, $j_0$, the initial angle, and both archived target components. The mass change is essential: changing length and time alone would not preserve the same canonical-momentum state. Training includes the identity and four fixed dimensional rescalings. These are a similarity-structured sampling prior.

### A.5 Gray–Scott reaction–diffusion

The qualified source is the two-dimensional Gray–Scott release in The Well (Ohana et al., 2024). The fields $A(x, y, t)$ and $B(x, y, t)$ are chemical concentrations. With diffusivities $D_A, D_B$, autocatalytic rate $\kappa$, feed rate $f$, and removal rate $k$, the source-linked generator implements

$$\partial_t A = D_A \nabla^2 A - \kappa A B^2 + f(1 - A), \tag{12}$$

$$\partial_t B = D_B \nabla^2 B + \kappa A B^2 - (f + k)B. \tag{13}$$

Diffusion spreads concentrations, the autocatalytic reaction converts $A$ to $B$, and feed/removal exchange material with the surroundings. Autocatalysis is positive in the $B$ equation. The source overview displays the opposite sign. The source generator uses the positive sign adopted here, supported by the saved finite-cadence residual check. The archive-conversion code was unavailable, so the coordinate conversion remains a source qualification. These checks do not bound solver discretization error; the full source audit is in the evidence package.

For positive rates and lengths, possible normalized controls include $\kappa t$, $f/\kappa$, $k/\kappa$, $D_A/(\kappa L^2)$ and $D_B/(\kappa L^2)$, together with aspect ratio and normalized initial concentrations. The complete action used here is

$$(x', y', L_x', L_y') = a(x, y, L_x, L_y), \quad t' = bt, \quad D_A' = \frac{a^2}{b}D_A, \quad D_B' = \frac{a^2}{b}D_B, \tag{14}$$

$$(\kappa', f', k') = \frac{1}{b}(\kappa, f, k), \qquad (A', B') = (A, B). \tag{15}$$

Native $D_A = 2 \times 10^{-5}$, $D_B = 10^{-5}$, $\kappa = 1$, and $L_x = L_y = 2$ are fixed. Feed/removal cells vary, but the qualified local input chart supports one physical elapsed-time direction and two similarity directions. We therefore assess elapsed-time dependence and equivalent descriptions, rather than independent recovery of every possible chemical group. Native cell comparisons retain explicit initial-field matches; a shared row number or random-seed label is insufficient evidence that the initial fields agree.

### A.6 Compressible flow

The compressible source is the two-dimensional periodic PDEBench CFD archive (Takamoto et al., 2022). Four source files cross nominal Mach labels $0.1$ and $1$ with viscosity coefficients $0.01$ and $0.1$. Analysis uses measured Mach values rather than substituting those filename labels. The source uses density-weighted, kinematic-like viscosity coefficients. Writing its coefficient as $\eta_{\mathrm{v}}$ to distinguish it from the sensitivity fraction $\eta$, the physical controls are

$$\mathrm{Ma} = \frac{U_{\mathrm{ref}}}{c_{\mathrm{ref}}}, \qquad \mathrm{Re} = \frac{U_{\mathrm{ref}} L}{\eta_{\mathrm{v}}}, \qquad \tau = \frac{t U_{\mathrm{ref}}}{L}. \tag{16}$$

Here $U_{\mathrm{ref}}$ is obtained from the archived initial speed and $c_{\mathrm{ref}}$ from reference density, pressure, and the source configuration's $\gamma = 5/3$. The bulk/shear coefficient ratio is fixed at $\zeta/\eta_{\mathrm{v}} = 1$. No additional density factor or Burgers coefficient conversion is inserted into Reynolds number. The source implementation uses direction-split diffusion; it is not replaced by a different continuum stress convention.

The source grid is $128 \times 128$ with channels density, pressure, and two velocity components. The four future frames are at $t = 0.05, 0.1, 0.5, 1.0$, plus the initial frame. A trailing coordinate at $1.05$ has no corresponding archived field and is excluded.

The complete action is $(x, y, L_x, L_y, t, \eta_v, \zeta) \mapsto a(x, y, L_x, L_y, t, \eta_v, \zeta)$, with initial and future $(\rho, p, v_x, v_y)$ unchanged at corresponding relative locations. It preserves the normalized state. Training uses seven logarithmically spaced factors from $2/3$ to $1.5$. The derivative analysis retains the four registered input-control directions and their three-dimensional physical span. Finite comparisons retain the registered same-Mach viscosity and matched-Reynolds Mach panels, both orientations, and all predefined steps. The finite action's log step $0.02$ should not be confused with displacement along a unit-normalized derivative direction.

### A.7 Acoustic scattering

Acoustics uses The Well's CC-BY-4.0 `acoustic_scattering_discontinuous` archive (Ohana et al., 2024), retaining pressure and both velocity components on the actual $256 \times 256$ cell-centered domain $[-1, 1]^2$, with WALL lower and OPEN upper boundaries. The pressure and velocity satisfy $\partial_t p = -K \nabla \cdot v$ and $\rho \partial_t v = -\nabla p$. Its complete amplitude-preserving action is

$$(x, y, L)' = a(x, y, L), \quad t' = bt, \quad K' = \frac{a}{b} K, \quad \rho' = \frac{b}{a} \rho, \quad (p, v)' = (p, v).$$

It preserves geometry, boundary types, normalized initial fields, acoustic travel progress and impedance. The clock is $c_{\mathrm{ref}} t / L_{\mathrm{ref}}$, where $c_{\mathrm{ref}} = N^{-1} \sum_{j=1}^{N} \sqrt{K_j / \rho_j}$ is the arithmetic mean of the pointwise speeds over the uniform grid and $L_{\mathrm{ref}} = \sqrt{L_x L_y}$. It is not $\sqrt{\overline{K} / \overline{\rho}}$. The supported three-dimensional chart spans two similarity directions and one time direction. Impedance variation is outside this chart; initial fields and material layouts vary together across native parents. No independent material-regime claim follows.

### A.8 Harmonic Schrödinger evolution

The harmonic Schrödinger source is the MIT-licensed archive (ajthor, 2025). Three official-training shards and one official-test shard supply 400 unique initial states, 41 archived times $0, 0.05, \ldots, 2$, and 256 periodic points $x_j = -10 + 20j/256$. The publisher's implemented evolution has the opposite Hamiltonian sign to its written equation. We complex-conjugate every archived initial field and trajectory, retaining the originals, to obtain

$$i\hbar \psi_t = -\frac{\hbar^2}{2m} \psi_{xx} + V_h \psi.$$

$V_h$ is the actual archived periodic spectral potential, not an ideal sampled quadratic; their maximum difference is $0.001526$. Derivative checks support the conjugated convention at archived cadence, without bounding continuum solver error. All reference fields are archived trajectories.

The complete Schrödinger action is

$$(x, L)' = a(x, L), \quad t' = bt, \quad m' = cm, \quad \hbar' = \frac{ca^2}{b} \hbar, \quad \omega' = \frac{\omega}{b},$$

$$\psi' = a^{-1/2} \psi, \quad V_h' = \frac{ca^2}{b^2} V_h.$$

This preserves quantum norm, $\hbar / (m \omega L^2)$ and $\omega t$. Native $m = \omega = \hbar = 1$, $L = 20$ are fixed. The supported four-dimensional chart spans three similarity directions and time, with only one varying physical direction. Clock removal leaves zero additional supported physical rank in both systems. Real/imaginary state distances are global-phase sensitive; no gauge alignment is used.

# B EXPERIMENTAL DESIGN AND EVIDENCE POPULATIONS

The publicly available simulation datasets are PDEBench for advection, Burgers and compressible flow, the Hamiltonian Dynamics Suite for pendulum mechanics, The Well for Gray–Scott and acoustics, and the harmonic Schrödinger release (Takamoto et al., 2022; DeepMind, 2021; Ohana et al., 2024; ajthor, 2025). Complete actions from Appendix A generate dimensional realizations of a normalized initial problem. Predictors receive dimensional inputs and supervised state losses, without dimensionless channels or representation targets. Normalization uses training data only. The similarity-structured sampling prior is known; no same-marginal, uncoordinated-sampling ablation isolates its causal contribution.

## B.1 PREDICTOR CONFIGURATIONS

| System/model | Fits / inits | Blocks / width; setting | Grid | Hook | Updates / batch | Select |
|---|---|---|---|---|---|---|
| Advection FNO | 2/2 | 8 / 64; 24 modes | 128 | 3 | 6,000 / 32 | V |
| Advection CNN | 2/2 | 7 / 128; dilations 1:64 | 128 | 3 | 6,000 / 32 | V |
| Burgers FNO | 6/3 | 8 / 64; 24 modes | 128 | 3 | 18,000 / 32 | V |
| Burgers CNN | 6/3 | 7 / 128; dilations 1:64 | 128 | 3 | 18,000 / 32 | V |
| Pendulum MLP | 3/3 | 4 / 384 | 6 scalars | 1 | 18,000 / 32 | V |
| Pendulum Transformer | 3/3 | 4 / 64; 4 heads | 6 tokens | 1 | 18,000 / 32 | V |
| Gray–Scott CNN | 3/3 | 7 / 48; dilations 1:64 | 128×128 | 3 | 18,000 / 32 | V |
| Gray–Scott FNO | 3/3 | 4 / 32; 16×16 modes | 128×128 | 1 | 18,000 / 32 | V |
| Compressible FNO | 3/3 | 6 / 48; 24×24 modes | 128×128 | 2 | 18,000 / 8 | V |
| Compressible CNN | 3/3 | 7 / 64; dilations 1:64 | 128×128 | 3 | 18,000 / 8 | V |
| Acoustic FNO | 2/2 | 4 / 32; 16×16 modes | 256×256 | 1 | 4,000 / 32 | V |
| Schrödinger FNO | 2/2 | 4 / 48; 24 modes | 256 | 1 | 6,000 / 32 | V |

**Table A1: Predictor configurations used in the study.** Fits/inits gives the number of trained fits and distinct random initializations. V denotes checkpoint selection by lowest validation prediction error. Hook indices refer to zero-indexed structural-middle blocks, with the complete vector, token array, or spatial feature field retained. The 38 trained fits come from 32 distinct initializations; the six Burgers G2/G3 pairs share initial weights across training arms.

Depth counts operator, residual or Transformer blocks, excluding the input lift and output projection. CNN blocks use two stride-one, kernel-size-three convolutions with the stated dilation and periodic padding. The one-dimensional FNO/CNN output projection has width 128; the two-dimensional projection has width 64. Acoustic padding is 16 cells in each direction. Schrödinger uses a width-96 projection and periodic grid. The MLP has two width-384 linear layers per residual block. The Transformer uses six scalar tokens, feed-forward width 256, four attention heads, zero dropout and a width-64 output hidden layer. Field hooks retain every site and channel; mechanical hooks retain the full vector or token array.

All predictors use Adam at initial learning rate $10^{-3}$ with cosine decay across the full update budget. The final rate is $10^{-4}$ for advection, Burgers and compressible flow, and $10^{-5}$ for mechanics, chemistry and both wave families. Batch sizes are effective examples per update; chemistry and acoustics accumulate eight microbatches of four. Mechanics, chemistry and Schrödinger clip gradient norm at one. Checkpoints are selected by ordinary validation prediction error. The principal synthesis has 38 fits from 32 distinct random initializations; six Burgers pairs share weights across two training arms. Advection contains one development-selected and one fresh model per architecture. Acoustics contains one development-selected and one fresh FNO; the two quantum FNOs are fresh.

## B.2 SPLITS AND EVIDENCE POPULATIONS

All descendants, windows, regimes and rescalings of a verified initial condition retain its split. Common calibration inputs are distinct from representation evaluation cases, although calibration can use predictor-seen training parents. Table A2 separates local, finite-family, prediction and confirmation populations. Advection uses 64 training, 16 validation, 40 correspondence FIT, 16

SELECT and 16 exploratory parents. Eight reserved contexts are used only in the separately frozen coverage confirmation. Burgers uses 632 training initial conditions, with 40 FIT and 20 SELECT parents drawn from training; 24 reused validation parents support development evaluation and checkpoint selection. Twelve reserved contexts remain closed. The two- and three-regime arms each receive 18,000 updates at batch 32 and share initial weights and non-regime streams. Both normalize from endpoint-regime training fields; equal regime cycles allocate 288,000 examples per endpoint in G2 and 192,000 per regime in G3. G2 correspondence calibration includes only endpoint regimes.

Pendulum has 700 training, 100 validation, 100 exploratory and 100 reserved parents before its gravity-band holdout. The holdout removes 57, seven and ten from the first three roles, leaving 643 training and 90 ordinary exploratory parents. It uses different trajectories, so it does not isolate matched gravity interventions. Inputs are $(\theta_0, p_0, m, l, g, t)$ and outputs $(\theta, p)$. Training uses 17 queries spanning frames 16–400, with four withheld within-range queries and three later queries spanning frames 448–511. Local assays use eight parents, whereas recurrence uses 24 and prediction 90. Correspondence uses 80 FIT and 16 SELECT predictor-seen training parents.

Compressible flow has 160 training contexts, partitioned into 120 FIT and 40 SELECT, plus 24 validation, 56 exploratory and 16 reserved contexts. The principal evaluation has six models, two stages and 56 contexts, or 672 model-stage-context cases. Its scope was reduced after partial exploratory evaluation; it is not a confirmatory sample. Four regimes and four positive times stay within each context.

Gray–Scott has 48 training, eight validation, eight exploratory and eight reserved parents. Cell $(f, k) = (0.029, 0.057)$ is excluded from updates, normalization, selection and correspondence. Four source windows spanning frames 0–750 share the parent role. Four supervised offsets span 1–8, two interpolation offsets lie within that range, and offset 16 is longer; one archive step is ten time units. The local assay has three parents, ordinary prediction eight. The distinction is retained in every uncertainty statement.

Acoustic roles contain 180 training, 26 validation, 25 exploratory and 25 protected parents. Four FIT and two SELECT parents are predictor-seen; four separate exploratory parents supply local and dense-time evaluation. Schrödinger's 300 official-training parents are split into 240 training, 30 validation, 20 FIT and ten SELECT. Official-test parents remain separate: four enter local/representation assays and 16 enter prediction. Four FIT and two SELECT parents support correspondence. Protected mechanical, chemical, Burgers and compressible response reserves remain closed.

| System / assay | Fits (inits) | Parents | Time/regime/description domain | Role | Estimator | Uncertainty unit |
|---|---|---|---|---|---|---|
| Advection; local | 4 (4) | 16 | 3 speeds, 3 native queries | D/F; E | Pooled | Whole IC |
| Advection; coverage | 8 (4) | 16 | 52 equal-progress pairs; 35 response contrasts | F; E | Pooled; distances | Whole IC |
| Advection; confirmation | 8 (4) | 8 | Same frozen coverage contrasts; held-out $c = 4$ | C | Pooled; distances | Whole IC |
| Burgers; local/finite | 12 (6) | 24 | 3 viscosities, 4 times in $[.25, 1]$; 3 steps in $[.005, .08]$ | F; V | Pointwise; relative | Whole IC |
| Burgers; finite families | 6 (6) | 3 | G2; 3 regimes, 21 descriptions; 21 comparisons/model | V | RMS ratios | Descriptive |
| Pendulum; local/finite | 6 (6) | 8 | 8 native anchors; small/wider finite descriptions | F; E | Pointwise; relative | Seed/parent |
| Pendulum; prediction | 6 (6) | 90 | 17 supervised queries; 4 interpolation, 3 later queries | F; E | Error; $R^2$ | Whole parent |
| Pendulum; recurrence | 6 (6) | 24 | Frames 0–511; native recurrence 0–400; later separate | E | Rank; 5-NN | Parent; rank descriptive |
| Gray–Scott; local/finite | 6 (6) | 3 | 4 anchors, offsets spanning 4–8; small/wider descriptions | F; E | Pointwise; relative | Seed/parent; LOPO |
| Gray–Scott; prediction | 6 (6) | 8 | 5 regimes, 4 windows; 4 supervised offsets in 1–8; 2 within-range; late 16 | F; E | Error; $R^2$ | Whole parent |
| Compressible; local/finite and prediction | 6 (6) | 56 | 4 regimes, 4 times in $[.05, 1]$; 3 steps in $[.005, .08]$ | F; E | Pooled; relative; error | Context; 3 seed blocks |
| Acoustic; local/finite | 2 (2) | 4 | 4 native anchors; 2 small/2 wider descriptions | D/F; E | Pointwise; relative | Parent; descriptive |
| Acoustic; prediction | 2 (2) | 4 | Native frames 1–52; original 8-frame/5-description panel retained | E | TRAIN-scale MSE | Parent spread |
| Schrödinger; local/finite | 2 (2) | 4 | 4 native anchors; amplitude-covariant descriptions | F; E | Pointwise; relative | Seed/parent |
| Schrödinger; prediction | 2 (2) | 16 | All 40 positive frames; 11 supervised, 29 intermediate | F; E | Complex relative $L^2$ | Parent spread |

**Table A2: Evaluation populations for the primary analyses.** Rows identify the models, physical parents, query ranges, analysis role, estimator, and uncertainty unit used by each assay; models reused across rows should not be counted as additional fits. D/F denotes development-selected/fresh models, E exploratory evaluation, V reused validation, and C confirmation. Whole-parent resampling keeps repeated times and transformed descriptions within the same parent block.

# C Representation measurements and robustness

## C.1 Dependency spaces and estimators

Relative logarithms give $\log(\Pi/\Pi_*) = B^\top s$ for monomial groups. Projecting the exponent directions into the supported input chart and orthonormalizing them gives $Q$ and $P = QQ^\top$. No feature score selects this physical subspace. For $G = J_z^\top J_z$,

$$S_{\text{phys}} = \frac{\operatorname{tr}(PG)}{\operatorname{tr}(G)}.$$

Derivatives include every affected input channel, coordinate, coefficient and quadrature term. Burgers, mechanics, chemistry, acoustics and Schrödinger average pointwise fractions with their original equal-parent hierarchy. Advection and compressible flow instead pool physical and total energies before division. These estimators are not interchangeable.

Invertible recombinations of log groups preserve the dependency span. If $z = h(B^\top s)$ and $J_h$ has full column rank, $\ker J_z = \ker B^\top$; otherwise features can ignore another physical direction. Locally invertible feature changes preserve this exact kernel but need not preserve energy fractions, singular values or leading subspaces. Orthogonal feature transformations and uniform nonzero scaling preserve the native share.

Unresolved total responses, constant targets and zero denominators remain undefined. Local derivatives are checked against central differences at log steps $0.005, 0.02, 0.08$, with the smallest resolved step distinguishing numerical noise from larger-step curvature. The physical share and leading subspace require their respective signal/gap checks. Conditional intervals use 2,000 whole-parent resamples with predictors and maps fixed; paired seed/parent intervals are identified explicitly. Repeated times and descriptions remain inside their parent block.

## C.2 Physical specificity, clock removal, and feature scaling

The reference contains 2,048 fixed rank-$k$ projectors in a supported chart of dimension $d$, generated by isotropic Gaussian QR with fixed signs and seed 26092613. Each orientation is shared across stages and anchors before the original pointwise or pooled reduction. Rotational symmetry gives $\mathbb{E}P_{\text{random}} = (k/d)I$ and $\mathbb{E}S = k/d$. Reference percentiles condition on the observed derivative spectrum; they are not population $p$-values or confidence levels.

For the unit clock gradient $u$ with $Pu = u$, let $C = I - uu^\top$ and $P_{\text{extra}} = P - uu^\top$. The clock-removed share is

$$S_{\text{extra}} = \frac{\operatorname{tr}(P_{\text{extra}}G)}{\operatorname{tr}(CG)}, \qquad \mathbb{E}S_{\text{extra,random}} = \frac{k-1}{d-1}. \tag{17}$$

The original input-chart metric and aggregation are preserved. Residual energy must exceed the local numerical threshold and $10^{-10}\operatorname{tr}(G)$. Burgers uses $u = (-1, 0, 1)/\sqrt{2}$ in $(\log L, \log \nu, \log t)$ and extra direction $(1, -2, 1)/\sqrt{6}$. The reference expectations are $1/2$ for Burgers, $1/3$ for mechanics and $2/3$ for compressible flow. Advection, chemistry, acoustics and Schrödinger have zero additional supported physical rank. This removes an input direction, not a fitted feature-time response.

For an invertible feature transformation $A$ with condition number at most $K$,

$$L_K(S) = \frac{S}{S + K^2(1-S)} \le S_A \le U_K(S) = \frac{K^2 S}{K^2 S + 1 - S}. \tag{18}$$

Singular-value bounds on physical/similarity energies give these inequalities. Pointwise estimates average pointwise bounds; when only their mean is used, convexity of $L_K$ and concavity of $U_K$ give conservative outer bounds. Pooled estimates use the inequalities directly. A positive trained-lower minus initial-upper gap certifies a positive measured change even if transforms differ across stages. Overlap is inconclusive and does not exhibit a reversal. These deterministic bounds do not cover arbitrary-coordinate invariance.

Practical gain tests use 32 fixed permutations of equally spaced log gains at each $K \in \{1, 2, 4, 10\}$, geometric mean one, shared across stages and spatial sites/tokens. Channel-resolved derivative Grams

are required. Shared calibration normalization uses

$$D_c = \left[\tfrac{1}{2}(v_{\text{initial},c} + v_{\text{trained},c}) + \epsilon\right]^{-1/2}, \qquad \epsilon = 10^{-6} \times \text{mean positive variance}.$$

The stage-specific companion uses each stage's own variance. Examples and sites within examples receive equal calibration weight. Mean variance must exceed $64\epsilon_{32}^2$ times mean squared feature; zero-variance channels retain regularization. Mechanics, chemistry and the two wave families use their complete local populations. Other field tests use 64 predetermined anchors spanning all original parents. At most 128 paired calibration examples are used, except acoustics has 100. These subsets do not replace complete-population native endpoints. Measured gains and normalizations do not isolate a causal contribution of channel amplitudes.

### C.3 Finite equivalent descriptions

For a family of equivalent descriptions $\mathcal{O} = \{z_s\}$,

$$\mu_{\mathcal{O}} = |\mathcal{O}|^{-1} \sum_s z_s, \qquad R^2_{\mathcal{O}} = |\mathcal{O}|^{-1} \sum_s \|z_s - \mu_{\mathcal{O}}\|^2. \tag{19}$$

Figure 3d compares complete-native radii with physical separation across three parents:

$$R^2 = \tfrac{1}{3} \sum_p \tfrac{1}{2}(r^2_{p,q} + r^2_{p,q'}), \qquad D^2 = \tfrac{1}{3} \sum_p \|\mu_{p,q'} - \mu_{p,q}\|^2.$$

Each of six G2 models supplies nine adjacent-time and twelve matched-progress regime comparisons. The 126 after/before ratios retain 21 descriptions per family. Temporal rosters match; adjacent/extreme regimes share 14/seven descriptions, so regime comparisons do not isolate viscosity alone. Separation below $10^{-12}$ of the maximum relevant native distance is unresolved. These are within-model ratios in each model's feature units.

Anchor-relative distance is $\|z' - z\|/\|z\|$, used in Burgers/compressible local finite assays. Mechanics and chemistry use the symmetric convention

$$d_{\text{sym}} = \frac{2\|z' - z\|}{\|z'\| + \|z\|}. \tag{20}$$

Absolute differences and endpoint magnitudes remain distinct from relative distances; transformation ranges are not pooled. Small and wider mechanical/chemical actions are $(a, b) = (0.95, 1.05)$ and its swap, or $(0.7, 1.4)$ and its swap. Burgers/compressible finite actions use $h = 0.005, 0.02, 0.08$ in their complete similarity transformations. The finite triplet step and displacement along a unit derivative direction are different parameterizations.

The mechanical illustration retains the fixed MLP/parent, frames 0–400, native and two small descriptions, separate-stage unwhitened calibration PCA3 and the original camera. Retained variance is 86.3%/85.8% in calibration, not the displayed population. Equal-description predictions may share physical error: for errors $e_a, e_b$,

$$\tfrac{1}{2}(\|e_a\|^2 + \|e_b\|^2) = \left\|\tfrac{1}{2}(e_a + e_b)\right\|^2 + \left\|\tfrac{1}{2}(e_a - e_b)\right\|^2.$$

Thus finite covariance alone does not establish physical accuracy.

### C.4 Cross-model correspondence

Common calibration inputs supply complete hidden features to PCA8 and regularized canonical correlation analysis. Correspondence is fitted without physical labels and frozen before a physical readout trained in one model is transferred to the other. Both directions, initialization and separate raw-input readouts are retained. Predictor-seen calibration parents and evaluation populations remain distinct. Core assays select regularization using separate SELECT correspondence; acoustic and quantum assays fix ridge $0.001$. The exact grids and secondary layer sweeps remain in the evidence package. Those choices use correspondence quality, never physical-readout scores. Readouts and correspondence are not refitted on evaluation cases.

### C.5 Response directions under equivalent descriptions

For adjacent actual observations from the same parent, regime and description,

$$v_s(\tau_i) = \frac{h_s(\tau_{i+1}) - h_s(\tau_i)}{\tau_{i+1} - \tau_i}, \qquad A_{\mathrm{dir}}(v_a, v_b) = \frac{v_a^\top v_b}{\|v_a\|\|v_b\|}.$$

No difference crosses a missing-frame or identity boundary. Unresolved differences stay undefined: displacement must exceed 32 float32 epsilons times the larger endpoint norm for Burgers, or their sum for mechanics. All primary vectors resolve. These are complete-feature responses to supplied clocks, not autonomous dynamics; no smoothing is applied.

Burgers uses three FNOs, three reused validation parents and three viscosities; its 75 primary intervals begin at frames 25–99, within supervision. Small log rescalings are $\pm 0.005, \pm 0.02$, wider ones $\pm 0.08$. Mechanics uses three MLPs and 24 exploratory parents, with 400 intervals through frame 400, including early queries before its first supervised frame. Small/wider actions are those in Appendix C.3. Intervals and descriptions average within parent, then parents equally. Changed-progress controls use lag 25 in Burgers and 64 in mechanics; changed-regime controls match progress across viscosities. Table A3 separates those control populations from matched descriptions and retains every wider mechanical decline.

| Family / seed | Comparison | Direction cosine: before → after |
|---|---|---|
| Burgers FNO | Small descriptions | 0.99984–0.99994 → 0.99998 |
| Burgers FNO | Wider descriptions | 0.99512–0.99824 → 0.99936–0.99953 |
| Burgers FNO | Changed progress | 0.970–0.974 → 0.234–0.472 |
| Burgers FNO | Changed regime | 0.942–0.986 → 0.930–0.941 |
| Pendulum MLP | Small descriptions | 0.958–0.963 → 0.997–0.999 |
| Pendulum MLP 2101 | Wider descriptions | 0.517 → 0.463 |
| Pendulum MLP 2102 | Wider descriptions | 0.495 → 0.414 |
| Pendulum MLP 2103 | Wider descriptions | 0.505 → 0.496 |
| Pendulum MLP | Changed progress | 0.968–0.970 → −0.057–0.053 |

**Table A3: Response-direction agreement under equivalent descriptions and physical changes.** Values are equal-parent mean direction cosines before and after training. Ranges summarize three model-specific means, while the wider pendulum results are shown by seed because all three decline. Changed-progress and changed-regime controls use physically different query pairs from the equivalent-description comparisons; the reported ranges are descriptive rather than uncertainty intervals.

### C.6 Recurrence

Pairwise archived states at their recorded times define physical recurrence (Eckmann et al., 1987). Mechanics uses unwrapped angle and canonical momentum under fixed output normalization; hidden features keep their native units. Excluding the diagonal and five adjacent archive steps, the tests compare distance ranks and five-neighbor overlap, with stable tie ordering. Initialization, inputs, clock-only and constant controls retain their own geometry. Native frames 0–400 and later frames 401–511 remain separate. All predictions are direct queries; no phase alignment or time warping is applied.

# D PRIMARY REPRESENTATION RESULTS

## D.1 ALL-SYSTEM SPECIFICITY AND ROBUSTNESS

Table A4 summarizes existing model-level results by family. It displays ranges, not a new pooled physics score or a count of statistically significant models. Figure 2 retains every fit. The native populations are 144 advection anchors from 16 parents, 288 Burgers anchors from 24 parents, eight mechanical anchors from eight parents, four chemical anchors from three parents, 896 compressible anchors from 56 contexts, and 16 anchors from four parents in each wave system. Alternative-direction references and channel-scaling controls were specified after the primary assays; they do not make the whole program one preregistered experiment.

| System/model | Fits | Native $\Delta S$ | Percentile | $S_{\mathrm{extra}}$ | Min. gain $\Delta S$ | $K = 2$ |
|---|---|---|---|---|---|---|
| Advection FNO | 2 | 0.578–0.746 | 99.8–100.0 | – | 0.543 | 2/2 |
| Advection CNN | 2 | 0.633–0.683 | 100.0 | – | 0.447 | 2/2 |
| Burgers FNO | 6 | 0.278–0.383 | 99.8–100.0 | 0.854–0.941 | 0.245 | 6/6 |
| Burgers CNN | 6 | 0.289–0.337 | 100.0–100.0 | 0.916–0.942 | 0.232 | 6/6 |
| Pendulum MLP | 3 | 0.471–0.546 | 100.0 | 0.954–0.979 | 0.424 | 3/3 |
| Pendulum Transformer | 3 | 0.102–0.200 | 77.5–88.7 | 0.542–0.630 | 0.064 | 0/3 |
| Gray–Scott CNN | 3 | 0.478–0.631 | 98.9–99.3 | – | 0.336 | 2/3 |
| Gray–Scott FNO | 3 | 0.537–0.688 | 97.9–99.1 | – | 0.447 | 2/3 |
| Compressible FNO | 3 | 0.288–0.339 | 99.6–100.0 | 0.973–0.981 | 0.284 | 3/3 |
| Compressible CNN | 3 | 0.192–0.219 | 100.0 | 0.991–0.992 | 0.279 | 3/3 |
| Acoustic FNO | 2 | 0.341–0.437 | 95.9–96.9 | – | 0.285 | 0/2 |
| Schrödinger FNO | 2 | 0.263–0.348 | 82.4–88.0 | – | 0.208 | 0/2 |

**Table A4: Physical specificity and feature-scaling robustness by model family.** Native $\Delta S$ reports the training-induced change in physical share, percentile compares trained representations with random subspaces of the same dimension, and $S_{\mathrm{extra}}$ gives the trained physical share after removing elapsed progress where additional physical rank remains. The gain column reports the smallest change across the measured channel-rescaling tests, and the final column counts models with a positive conservative $K = 2$ conditioning certificate. Ranges summarize saved model-level results without pooling systems.

Clock-removed Burgers shares increase from $0.343$–$0.523$ to $0.854$–$0.942$ against expectation $0.5$. Its trained FNO reference percentiles are $88.3$–$97.4$ and CNN percentiles $95.9$–$97.9$, so the additional direction is not uniformly outside the central 95% reference range. Mechanical and compressible residual dependence remains substantial, with weaker mechanical Transformer organization. The acoustic and quantum comparisons test a single progress direction and have lower reference ranks; no additional rank remains after clock removal.

All 38 fits have positive paired changes under the 128 measured channel-gain assignments and both calibration normalizations. The deterministic $K = 2$ bound certifies 29/38, including every advection/Burgers/compressible model and each mechanical MLP, but none of the Transformers or wave/quantum models. Chemistry certifies two of three models per architecture. For MLP, the largest initial upper bound is $0.806852$ and smallest trained lower bound $0.931029$. Overlapping bounds are inconclusive, not evidence of reversal. The fixed $K = (1, 1.5, 2, 4, 10)$ certificate counts are $(38, 33, 29, 1, 0)$; tighter pointwise bounds give the same counts. Positive measured transforms do not imply invariance under every invertible encoding.

## D.2 MECHANICS

Table A5 retains the six seeds because the MLP/Transformer difference limits architecture-general interpretation. Shares average local fractions within the eight-parent hierarchy; temporal response error compares predicted and true changes under the same state normalization.

The MLP share gain is $0.49702$, with 95% paired seed/whole-parent interval $[0.41905, 0.58573]$; its response-error change is $-0.78723$, interval $[-0.91241, -0.63188]$. The Transformer gain is $0.14123$, interval $[0.07137, 0.21099]$; its response-error change is $-0.13981$, interval

| Architecture | Seed | $S$: before → after | Extra share | Native $E_\Delta$ | Small distance: before → after |
|---|---|---|---|---|---|
| MLP | 2101 | 0.5108→0.9818 | 0.9536 | 0.2187 | 0.3851→0.0500 |
| MLP | 2102 | 0.5097→0.9842 | 0.9609 | 0.2724 | 0.3950→0.0502 |
| MLP | 2103 | 0.4421→0.9878 | 0.9793 | 0.1751 | 0.3684→0.0416 |
| Transformer | 2101 | 0.5029→0.6047 | 0.5419 | 0.8215 | 0.1805→0.3105 |
| Transformer | 2102 | 0.4779→0.6777 | 0.6300 | 0.7909 | 0.2030→0.2623 |
| Transformer | 2103 | 0.4941→0.6162 | 0.5495 | 0.8837 | 0.2138→0.3099 |

**Table A5: Pendulum representation and physical-response results by seed.** Physical share, clock-removed share, temporal-response error, and finite equivalent-description distance are reported separately for the three MLP and three Transformer fits. The table retains seed-level results because the architecture difference is substantial. Local representation measurements use eight exploratory parents; later-time and ordinary prediction populations are evaluated separately.

$[-0.39425, 0.06347]$. The 2,000 resamples pair stages and resample seed and parent blocks. Three seeds and eight parents limit uncertainty assessment.

Later exploratory queries retain MLP share $0.99698$ but response error $0.76993$; the Transformer has $0.59771$ and $0.98840$. These use their separate late population. Ordinary prediction uses 90 parents, not eight. High unwrapped-angle $R^2$ can coexist with phase error because angle has large dynamic range. Neither architecture reproduces source-level conservation accuracy. Wider finite transformations leave substantial hidden movement; small-description MLP organization is stronger than Transformer organization (Appendix F).

### D.3 Compressible flow

Native and clock-removed shares are high in both architectures, while normalized physical-response errors remain large (Table A6). Pooled sensitivity fractions differ from mean-per-contrast response errors; one cannot assign the latter's uncertainty to the former. Mean resolved input rank remains four, so a large physical energy fraction does not mean an exact similarity kernel.

| Family | Trained $S$ | Clock-removed $S$ | Viscosity response: $p$ | Mach response: $v_x$ |
|---|---|---|---|---|
| FNO | 0.988–0.991 | 0.973–0.981 | 2.6910–3.0281 | 1.2003–1.2522 |
| CNN | 0.994–0.995 | 0.991–0.992 | 2.3046–2.9170 | 0.9719–1.0310 |

**Table A6: Compressible-flow physical selectivity and response error.** Ranges summarize the three FNO and three CNN predictors. Native and clock-removed shares use pooled sensitivity energies across 56 exploratory contexts. Response errors measure predicted physical changes for viscosity and Mach contrasts and use their own true-change denominators, so they should not be interpreted as snapshot-prediction errors.

The separate mean-per-anchor similarity fractions are $0.03592$ for FNO and $0.01366$ for CNN, with 95% intervals $[0.02875, 0.04068]$ and $[0.01057, 0.01767]$. Their paired difference is $0.02226$, interval $[0.01728, 0.02917]$, from 2,000 crossed resamples of three paired seed blocks and 56 contexts. These do not replace pooled physical fractions. The leading physical subspace remains poorly aligned despite the energy emphasis. At $t = 1$, every $v_x$ skill relative to the training-mean baseline is negative: $-0.275$ to $-0.252$ for FNOs, $-0.135$ to $-0.037$ for CNNs. Skill is $1 - \mathrm{SSE}_{\mathrm{model}}/\mathrm{SSE}_{\mathrm{baseline}}$.

### D.4 Gray–Scott reaction–diffusion

The three-parent local share rises from $0.261805$ to $0.896197$ for FNO and $0.213532$ to $0.792488$ for CNN. Paired seed/whole-parent change intervals are $[0.53747, 0.69835]$ and $[0.47793, 0.65306]$. Table A7 keeps seed detail and the separate eight-parent prediction estimator. Whole-parent omission leaves every local share change positive: FNO omission ranges by seed are $[0.489, 0.562]$, $[0.664, 0.690]$, $[0.676, 0.701]$; CNN ranges are $[0.429, 0.505]$, $[0.618, 0.647]$, $[0.607, 0.661]$. These 18 sensitivity values add no independent replication or confidence interval. Small finite relative distances decrease, while FNO absolute differences grow and CNN absolute differences shrink.

Short-horizon source-anchored errors are small, showing learned accumulated change beyond persistence. The excluded feed/removal cell also has useful short-offset predictions, but at offset 16 every held-out-cell channel $R^2$ is negative across all six models. This is longer direct-query prediction beyond supervised offset eight. Local two-future response errors instead average $2.27768$ (FNO) and $1.82609$ (CNN); change intervals $[-0.16948, 2.6448]$ and $[-0.48433, 2.46000]$ include zero. They do not establish resolved deterioration, and their small-increment denominator differs from accumulated source-window evolution.

| Family | Seed | $S$: before → after | Small distance: before → after | Source error $A/B$ |
|---|---|---|---|---|
| FNO | 2201 | 0.34519→0.88266 | 0.22792→0.02483 | 0.10211 / 0.15751 |
| FNO | 2202 | 0.21447→0.89184 | 0.21779→0.02798 | 0.10174 / 0.15663 |
| FNO | 2203 | 0.22576→0.91409 | 0.18461→0.02552 | 0.09918 / 0.15041 |
| CNN | 2201 | 0.30510→0.78303 | 0.40955→0.01087 | 0.01885 / 0.02883 |
| CNN | 2202 | 0.14871→0.77681 | 0.47716→0.01027 | 0.01911 / 0.02923 |
| CNN | 2203 | 0.18679→0.81763 | 0.41566→0.01088 | 0.01921 / 0.02968 |

**Table A7: Gray–Scott representation and prediction results by seed.** Physical share and finite equivalent-description distance use the three-parent local representation assay, while source-anchored prediction errors use eight exploratory parents. All six models show increased physical share and reduced small-description distance. Longer held-out-cell predictions at offset 16 are reported separately in the text and are not combined with the local representation measurements.

### D.5 Acoustic and Schrödinger evolution

Both systems have two FNOs and one supported physical direction. Their gains concern normalized progress, not independently varied impedance or quantum regime controls. Table A8 combines local dependence, common-protocol specificity, shared normalization and prediction, retaining distinct populations.

| System | FNO | $S$: before → after | Percentile | Shared-normalized $S$ | Supervised | Intermediate |
|---|---|---|---|---|---|---|
| Acoustic | 1 | 0.348→0.785 | 96.9 | 0.330→0.759 | 0.271467 | 0.767880 |
| Acoustic | 2 | 0.401→0.742 | 95.9 | 0.417→0.716 | 0.294535 | 0.807160 |
| Schrödinger | 1 | 0.140→0.403 | 82.4 | 0.121→0.403 | 0.180968 | 0.202041 |
| Schrödinger | 2 | 0.069→0.417 | 88.0 | 0.075→0.377 | 0.179615 | 0.198055 |

**Table A8: Acoustic and Schrödinger representation and prediction results.** Each system contains two FNOs. Physical share, random-subspace percentile, and shared-normalized share use the local four-parent representation populations; prediction columns report the dense-time errors shown in Figure 4. The acoustic and Schrödinger charts each contain one varying physical direction, so no additional physical rank remains after removing elapsed progress.

Quantum local-share gain averages $0.30515$, with the saved exploratory two-seed/whole-parent 95% bootstrap interval $[0.25572, 0.35103]$ from 2,000 draws. This small-sample interval is conditional, not confirmatory. The standardized-input share is $0.05980$. Acoustic intermediate/late errors in the original two-frame assays worsen from initialization; the broader dense-time means above use the same predictors on different timestamp populations.

Acoustic small-description relative distances decrease from $0.21$–$0.29$ to $0.030$–$0.037$ while absolute differences increase. Its small/wider scales match the mechanical actions in Appendix C.3. Quantum actions use small $(a, b, c) = (0.95, 1.05, 1.05)$ and $(1.05, 0.95, 0.95)$, and wider $(0.8, 1.25, 1.1)$ and $(1.25, 0.8, 0.9)$. Output comparisons remove the known amplitude factor; covariance residuals remain nonzero and increase despite better prediction. Nearly constant initial outputs can agree across descriptions without solving the task.

# E PHYSICAL BEHAVIOR AND CROSS-MODEL GENERALIZATION

## E.1 PHYSICAL EVOLUTION

Figure 4 compares direct queries with archived physical fields. Schrödinger uses 16 official-test initial fields, both FNOs and all 40 positive times $0.05, 0.10, \ldots, 2.0$. Its complete complex-field error is

$$e_i^{\mathrm{Sch}}(t) = \left[ \frac{\sum_x |\widehat{\psi}_i(x,t) - \psi_i(x,t)|^2}{\sum_x |\psi_i(x,t)|^2} \right]^{1/2}. \tag{21}$$

Both real and imaginary errors enter the numerator; uniform quadrature cancels. There is no phase alignment, density substitution or per-frame norm adjustment. The illustration shows the real component of the fixed first test parent and first FNO. Native supervision is $0.05, 0.2, 0.4, \ldots, 2.0$; the remaining 29 times are intermediate.

Acoustics uses four exploratory initial fields, both FNOs and native frames 1–52, whose timestamps span $0.0198019799$–$1.0297030210$. Supervised frames are 1,2,26,52; the other 48 are intermediate. The displayed pressure is the fixed first parent/FNO at frame 13 ($t = 0.2574257553$). Its full-state error is

$$e_i^{\mathrm{ac}}(t) = \frac{1}{3N_xN_y} \sum_{x,y,c} \left[ \frac{\widehat{u}_{i,c} - u_{i,c}}{s_c} \right]^2, \quad c \in \{p, v_x, v_y\}, \tag{22}$$

with fixed training scales $s_c = (0.403519515, 0.057309474, 0.056903882)$. This metric includes both velocities although the displayed field is pressure. Its units and normalization differ from complex relative $L^2$.

Lines are equal whole-parent means at each time; bands are parent min–max spread, not confidence intervals across seeds. Persistence holds the actual initial full state fixed under the same metric. Transformed descriptions add no native supervision nodes. Table A9 gives the complete native dense-time populations.

Quantum persistence averages $1.289506$ over all 40 times. Acoustic persistence averages $1.350190$ at the 48 intermediate frames. The two acoustic mean-error maxima are $1.247071$ at frame 12 and $1.339573$ at frame 11. The original two interpolation frames have errors $1.007687/1.054198$, and later frames $0.888977/0.963060$; those narrower populations do not replace dense-time means.

For the main-text response error $E_\Delta$, quantum native endpoints are frame pairs $1 \to 4$, $12 \to 16$, $24 \to 28$, $36 \to 40$; acoustic pairs are $1 \to 2$, $2 \to 13$, $26 \to 39$, $52 \to 77$. Acoustic thus mixes supervised, intermediate and later changes. Errors are approximately $0.91$ in acoustics and $0.336$–$0.347$ in quantum evolution. These keep their original true-increment denominators and populations; they are not recursive rollout errors.

## E.2 GENERALIZATION UNDER CHANGED PHYSICAL COVERAGE

Table A10 contrasts existing interventions; it does not compare systems under one treatment or isolate nonlinearity. G2 withholds the middle Burgers regime, while G3 uses all three. Advection enlarges dimensional descriptions at fixed physical task and target stream. Its selected domain was defined descriptively from metadata after the original broad endpoint, then frozen before eight-context confirmation. Exploratory and confirmation outcomes remain separate.

Across the complete description bank, G2 direct errors are $2.583$–$3.073\%$ and G3 errors $2.001$–$2.302\%$. G2 improves over its own endpoint-query average in six point estimates and five conditional intervals, but no full-bank point estimate beats the privileged true-endpoint average. Beating one averaging rule does not exclude other interpolation strategies or demonstrate recovery of the PDE. The middle-only $\log \mathrm{Re}$ target is constant, so its $R^2$ is undefined.

Advection selected-domain improvements reproduce in all four context-pooled model pairs, while one enriched CNN context still has $114.7\%$ error. Complementary reserve errors remain $99.43$–$100.48\%$. Enrichment includes speed 4; evaluated tuples are absent from training but their marginals lie within enriched ranges. This tests new combinations, not unseen-speed extrapolation. The original 16-parent exploratory population and later complete-circuit direct-query failures are separate from the confirmation primary.

**Table A9: Dense-query prediction errors for the two wave systems.** Each entry is an equal parent–time mean. Schrödinger uses 16 test parents over 40 positive times, including 29 intermediate timestamps; acoustics uses four parents over 52 frames, including 48 intermediate timestamps. The two systems use different error metrics and should not be compared numerically across rows.

| System | Predictor | All times | Supervised | Intermediate |
|---|---|---|---|---|
| Schrödinger | FNO 1 | 0.196246 | 0.180968 | 0.202041 |
| | FNO 2 | 0.192984 | 0.179615 | 0.198055 |
| Acoustics | FNO 1 | 0.729694 | 0.271467 | 0.767880 |
| | FNO 2 | 0.767727 | 0.294535 | 0.807160 |

| System | Intervention | Primary comparison / outcome | Population / scope |
|---|---|---|---|
| Burgers | Middle regime withheld | G2 direct $2.4225$–$3.0253\%$ vs own endpoint average $3.1012$–$3.5757\%$; all six paired direct errors lower | Canonical middle regime; reused validation |
| Advection | Broader equivalent descriptions | Response $83.06$–$121.37\% \rightarrow 34.01$–$47.38\%$; every pair improves | Four pairs; eight-context selected-domain confirmation |
| Burgers | Scale enrichment | Response $22.130$–$28.402\% \rightarrow 26.853$–$32.693\%$; all four paired errors worsen | Full validation IC/regime/time roster; distinct from regime holdout |

**Table A10: Generalization under changed physical coverage.** The table summarizes three distinct interventions: withholding the middle Burgers regime, broadening equivalent-description coverage in advection, and adding Burgers scale coverage. Ranges report existing model-specific endpoints and paired directions rather than cross-system averages or confidence intervals. Each intervention retains its original evaluation population and should be interpreted separately.

Nonlinear Burgers enrichment applies $(x, t, L, \nu) \mapsto k(x, t, L, \nu)$, $k \in \{0.5, 1, 2\}$ with $u, U$ fixed. Validation includes withheld $k = 0.85, 1.15$. Canonical field error worsens in every pair. Three of four paired regime-response intervals exclude zero on the unfavorable side; changes in equivalent-output discrepancy have intervals crossing zero. Relative hidden stability alone does not establish improved physical accuracy.

### E.3 Cross-model correspondence and transfer

Correspondence and physical-readout transfer follow Appendix C.4. Table A11 separates system, quantity and direction. Negative transferred $R^2$ is retained. Pairs share model identities, so their directional ranges are descriptive and not independent replications.

Mechanical signed-$j_0$ transfer spans $[-23.02843, 0.67111]$ with median $0.49430$ despite its strongly negative mean. Time spans $[-0.43306, 0.27470]$ and angle $[-1.52241, 0.67257]$. These heterogeneous results do not establish that all nonlinear correspondences fail. Burgers rescaled normalized-time transfer can improve despite reference-description decline; all 30 fresh-cohort directions improve under far scale extrapolation, a distinct domain.

Acoustic shared clock decoding is already $0.924$ at initialization and $0.925$ after training. Native CKA rises $0.888 \rightarrow 0.987$ for acoustics and $0.867 \rightarrow 0.997$ for quantum evolution; raw-time readouts are $0.793/0.824$. Parent shifts preserve time schedules, while time reversal disrupts transfer. These controls identify shared temporal information without establishing common material or quantum-state coordinates.

### E.4 Recurrence and decoded physical state

Across three trained MLPs, native hidden-to-physical distance-rank agreement is $0.016$–$0.041$, near clock-only $0.023$, while decoded outputs achieve $0.877$–$0.893$. Hidden five-neighbor overlap is about $2.2\%$, versus $26.4$–$31.0\%$ for outputs. Transformer output ranks are $0.100$–$0.148$. At later

| System / pairing | Physical quantity | Trained transfer $R^2$ | Control / interpretation |
|---|---|---|---|
| Burgers G2; all directed pairs | $\log \mathrm{Re}$; $\log \tau$ | $[-0.385, 0.876]$; $[0.731, 0.895]$ | No reference-domain direction improves over initialization |
| Burgers G3; all directed pairs | $\log \mathrm{Re}$; $\log \tau$ | $[0.739, 0.921]$; $[0.771, 0.891]$ | Some rescaled transfers improve; reference declines persist |
| Pendulum; MLP $\leftrightarrow$ Transformer | $\tau_{\mathrm{m}}, j_0, \theta_0$ | Means $-0.01227$; $-3.22318$; $-0.05190$ | Raw-input readout $(0.94004, 0.84802, \simeq 1)$ |
| Acoustic; two FNOs | Log progress | Matched $0.911/0.932$; parent-shift $0.908/0.929$ | Before matched $0.469/0.574$; time-reversed $-1.777/-1.427$ |
| Schrödinger; two FNOs | Log progress | Matched and parent-shift $0.963/0.965$ | Before $0.942/0.852$; time-reversed $-1.952/-2.103$ |

**Table A11: Cross-model transfer of physical variables.** The table summarizes frozen middle-layer correspondence and physical-readout transfer for Burgers, pendulum, acoustic, and Schrödinger models. Burgers ranges contain within- and cross-architecture directions, pendulum values summarize the overlapping MLP–Transformer pairs, and slash-separated wave values retain the two transfer directions. Raw-input readouts are controls for available information and are not transferred neural coordinates.

times, MLP output ranks fall to $0.432$–$0.587$, Transformer ranks to $0.036$–$0.039$. Each range uses the same 24 parents. Weak hidden Euclidean recurrence does not imply absence of all recurrent state information: the decoder can use a different geometry.

Burgers temporal-change medians on nine parent–regime cases are $0.133$ for FNO and $0.454$ for CNN under dense within-range queries, despite field errors $0.020/0.032$. Original and dense queries have different interval populations; their errors are not one matched interval estimator. This distinction complements the mechanical difference between native-feature distances and decoded physical state.

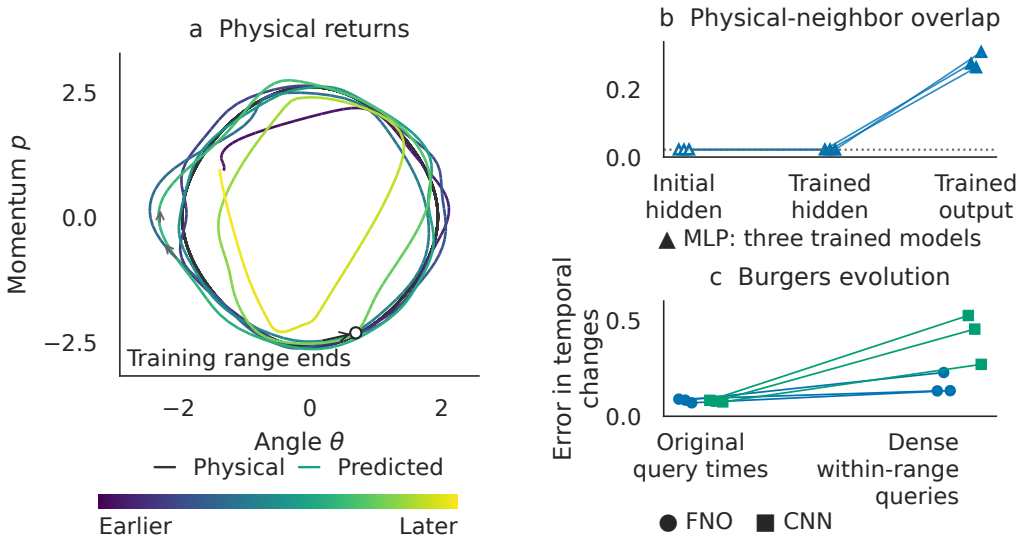



**Figure A1: Using learned physical variables.** (a) Physical and predicted pendulum trajectories for a fixed trained MLP under a small equivalent rescaling. The marker denotes the last supervised frame, 400; later direct-query predictions remain visible. (b) Mean physical-neighbor overlap for each MLP across 24 parents in the native time range. The dotted line is the supplied-clock control. Outputs recover physical recurrence much more strongly than middle-layer features. (c) Burgers temporal-change error for all three FNO and CNN models, reported as medians over nine parent–regime cases. Original and dense within-range queries use different interval populations and should not be compared as a matched estimator. All predictions are direct queries; Appendix E.4 reports Transformer and later-time recurrence results.

# F SUPPORTING FINITE GEOMETRY

## F.1 ARCHITECTURE ROUTES TO FINITE PHYSICAL ORGANIZATION

Figure 3 distinguishes relative physical separation from absolute description contraction. For three Burgers FNOs, small-description RMS spread rises from $0.0099$–$0.0117$ to $0.0780$–$0.0838$, while temporal separation rises $0.078$–$0.147 \rightarrow 20.70$–$22.77$ and regime separation $0.175$–$0.212 \rightarrow 7.20$–$10.53$. For CNNs, small-description spread instead contracts $0.341$–$0.404 \rightarrow 0.101$–$0.125$, while physical separation increases. Wider comparisons retain this qualitative difference. Every one of the 126 complete-native comparisons stays in Figure 3d; no projected width supplies those ratios.

The mechanical MLP's small-description spread-to-temporal-separation ratio falls $3.91$–$4.40 \rightarrow 0.088$–$0.123$. At wider descriptions, trained ratios are $3.23$–$5.56$; absolute spread contracts for one seed and grows for two. Transformer small-description ratios remain $2.85$–$3.32$, with absolute spread increasing in all seeds. Its temporal-response errors are $0.851$–$0.930$, versus MLP $0.272$–$0.320$, on this finite-family population. Those errors do not replace the eight-parent local increment assay. Native selectivity, finite geometry and accurate physical response are different measurements.

## F.2 COMPLETE SEVEN-SYSTEM QUALITATIVE PROJECTIONS

The montage retains every original fixed-default coordinate and sample identity. Main Figure 1 enlarges four examples; all seven systems remain in quantitative comparisons. Each before/after pair uses the same physical query roster. Physical progress colors the finished maps and does not select their coordinates.

The display populations have 204 advection queries, 2,700 Burgers queries, 2,880 mechanical queries, 300 chemical queries, 336 compressible queries and 300 queries each for acoustics and Schrödinger, per stage. Burgers includes frames 1–24 before its first supervised query; triangles identify these only in the full montage. Mechanics uses frames 0–400. Chemical display offsets extend through 32, beyond supervised offset eight. These are illustrative populations distinct from the local and prediction assays in Appendix B.

Fixed centered, unwhitened PCA up to rank 50 precedes t-SNE. Mechanics/Burgers/chemistry preprocess the fixed display population; the other four systems use FIT paths. The fixed t-SNE setting is perplexity 30, seed 260926, PCA initialization, automatic learning rate, 1,000 iterations and Barnes–Hut angle 0.5. Independently fitted axes and overall widths are incomparable. Complete settings and projection-quality checks are in the evidence package; no alternate setting supplies a main example. Additional nonlinear condition-geometry analyses were exploratory and heterogeneous; complete methods and outcomes are retained in the evidence package.

### F.2.1 CONTEXT-RESOLVED VIEWS OF FRAGMENTED PROJECTIONS

The raw population maps mix progress with differences among initial conditions and regimes. Figure A3 retains every query from Figure A2. Panels a,c repeat its saved initialization t-SNE maps for Gray–Scott and compressible flow. Panels b,d retain the trained PCA diagnostics: chemistry is centered separately within its 15 parent/feed–removal contexts; compressible trajectories are translated to their earliest responses. The two columns use different transformations and projections, so widths, distances and directions cannot be compared across columns.

Because absolute compressible-flow coordinates remain heterogeneous across contexts, we additionally visualize baseline-relative trajectories to test whether the direction of representational evolution becomes more consistent with training. For each stage, parent, regime and dimensional description, we subtract the earliest hidden response from all four saved queries. Seven descriptions remain separate paths, giving 84 trajectories per stage (12 parent/regime groups) and preserving all 336 identities. The existing PCA basis was fitted to the pooled before/after baseline-relative native vectors. Panel d retains the trained linear projections onto its first two axes, subtracting the projection of the zero vector from ordinary PCA scores. No individual trajectory is rotated, rescaled or aligned. The matched before/after PCA views remain in the evidence package.

The compressible PC1–2 capture 75.7% of pooled centered variance, but only 1.6% of before-stage and 66.0% of after-stage variance when each stage is centered separately. Thus this plane omits most

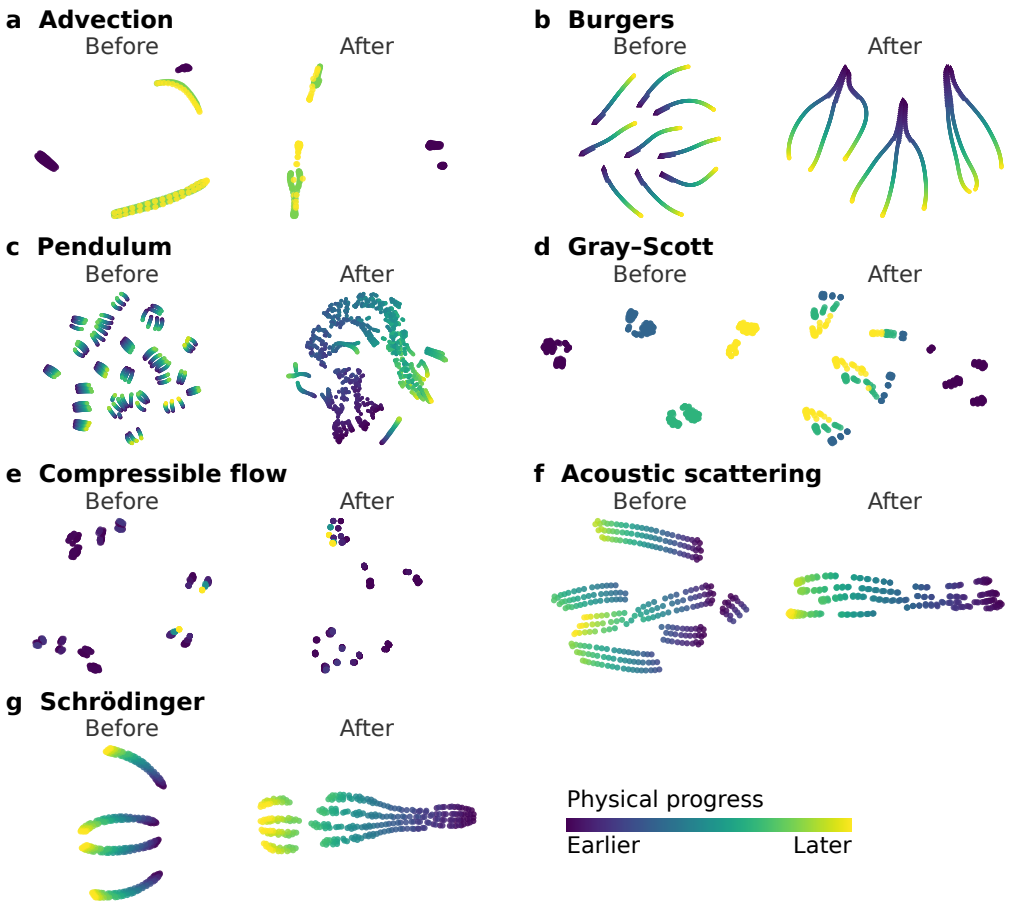



**Figure A2: Complete seven-system population views.** Matched hidden responses before and after training are shown for all seven physical systems and colored by physical progress. Each system and stage uses its own fixed projection, so axes and apparent widths are not directly comparable. Burgers triangles mark queries earlier than the first supervised query. Figure 1 enlarges four representative systems; all seven remain included in the quantitative analyses.

initial variation and cannot establish increased directional consistency from visual overlap. The fixed PCA3 inspection captures 83.2% pooled, 3.6% before and 75.9% after. Larger trained-stage variation dominates the joint fit. In panel d, lines connect saved queries, not evaluated intermediate states, and every path starts at the origin by construction. Chemistry's later queries still extend beyond supervision. These diagnostic views do not change any quantitative claim or establish invariant coordinates or accurate physical evolution.

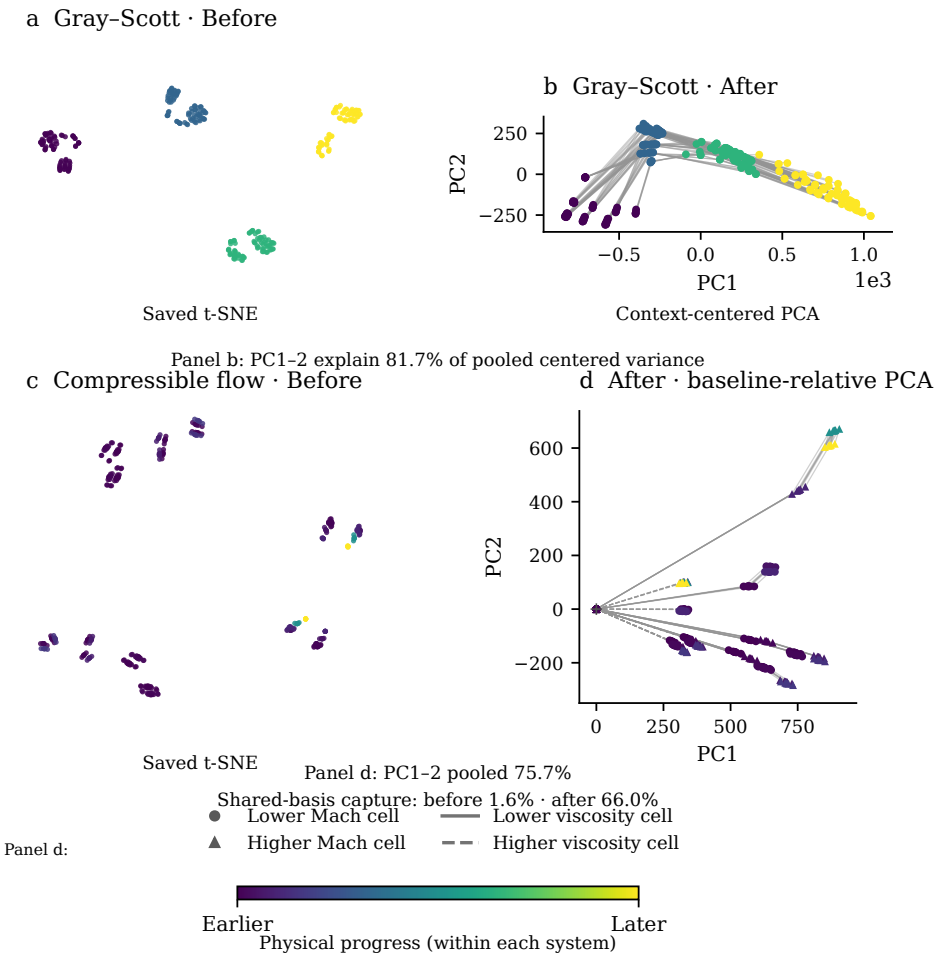



**Figure A3: Context-resolved Gray–Scott representations and baseline-relative compressible-flow evolution.** (a,c) Saved before-training t-SNE maps for Gray–Scott and compressible flow. (b) Trained Gray–Scott view after centering within each parent/feed–removal context. (d) Trained compressible-flow trajectories after subtracting each trajectory's earliest hidden response, projected onto a PCA basis fitted jointly to before/after responses. Color denotes physical progress; in d, marker shape distinguishes Mach cells and line style distinguishes viscosity cells. The shared PCA plane captures only 1.6% of before-stage variance but 66.0% after training. The columns use different projections, so they do not support a direct geometric before/after comparison. These panels are diagnostic views only; all quantitative claims use the original native-feature measurements.

# G NUMERICAL VALIDATION AND ANALYSIS SCOPE

## G.1 NUMERICAL VALIDATION AND SOURCE CORRECTIONS

The initial Burgers derivative extraction used TF32 arithmetic that failed systematic small-step checks. The corrected complete-population extraction disables cuDNN/matmul TF32 while preserving predictor weights, inputs, hooks, perturbations and correspondence. Across 864 selected middle anchors, mean small-step discrepancies are $0.0217$–$0.0573\%$ for similarity and $0.00480$–$0.00769\%/0.00670$–$0.01123\%$ for the physical directions. Larger steps include curvature. Original and corrected outputs remain in the evidence package; the corrected extraction supports the reported endpoints.

Acoustic and quantum predictors use FP32/complex64 with FP64 dimensional construction and reductions, with TF32 disabled. Their 128 local cases retain automatic derivatives, central finite differences, repeat and batch-context controls. Local signal and leading-gap checks resolve. Trained acoustic maximum small-step discrepancies are $0.000369/0.000502$; quantum minimum-step-discrepancy medians are $0.000142/0.000100$. Numerical resolution does not establish prediction competence.

The Burgers source coefficient conversion, Gray–Scott positive autocatalytic sign and quantum conjugation/Hamiltonian-sign correction are specified in Appendix A. The chemical coordinate-conversion qualification and the quantum archived spectral potential are retained there. These tests check implementation and archived-cadence consistency; they do not bound source-solver discretization error.

## G.2 ANALYSIS SCOPE

The primary claims use the predictor fits and populations in Appendix B. Repeated timestamps and transformed descriptions are not independent models; shared initializations are explicit. Exploratory, validation and confirmation populations remain distinct. The similarity-structured sampling prior is not causally isolated, and the mechanism of representation learning remains unresolved. Auxiliary and historical analyses add no confirmation to the headline claims. Negative outcomes directly limiting those claims remain reported: incomplete coordinate transfer, response errors despite selectivity, finite-range deterioration and incomplete feature-conditioning certificates.

Historical architecture screens, canceled branches, superseded assays and rejected auxiliary analyses are preserved in the accompanying evidence package and do not contribute additional confirmation to the final claims. Protected reserves remain closed except for the separately authorized frozen advection confirmation. The source-invalid logistic reaction–diffusion branch does not support the qualified Gray–Scott claims.