Supplementary Information: Predictive Structure and Realized Forecast Gain in Nonlinear Fields

Overview: evidence roles and reading map

This combined supplement preserves the detailed methods and results of the two component studies while assigning each its role in the present manuscript. Part A supplies the field representation diagnostics, transfer/horizon and molecular boundary cases, separate sensing experiments and numerical-certificate summary. Part B supplies the complete driven-lattice formation, observation, prospective-evaluation and statistical-proof record. A24 adds a measured-signal fixed-predictor history diagnostic; A25 adds the prospective 720-state field validation and common-output-space controls. Section identifiers are namespaced A and B; descriptor S14 retains its original physical meaning. References below are common to both parts. Configuration labels are identifiers, not additional experiments.

Main-text object	Detailed support	Evidence status
Known-mode ensemble, encoder/readers	A1, A2, A9, A14, A17, A19	Corrected cache; train-only fits; original splits
Same-field products and component errors	A11, A20, A23	Original exploratory diagnostic; separate common-reader control
Fixed-code readers and external cache	A23	45 configurations; nine frozen readers; previously examined external cache
Prospective fixed-code validation	A25	720 new states; one prespecified primary; common-output controls secondary
Historical-input dependence in measured records	A24	Fixed-model, post-hoc diagnostic; two previously examined records
Lattice codes and frozen predictors	B1–B5	Older training bank; fixed resolution/lookup before new outcomes
Information and score endpoints	B6, B8	Eight conditional lower bounds; unchanged multiplicity family
Horizon/transfer and molecular boundaries	A4, A13, A21	Their own units, losses and inferential scopes
Additional query/site sensing	A3, A5–A7, A10, A12, A16	Supporting application; corrected E3 primary unconfirmed
Finite-task coded-model guarantees	A8, A18; Supplementary Proofs SP0–SP4	Separate finite sets, pipelines and true-ODE contracts

Main-text object	Detailed support	Evidence status
Access and reproducibility	A15, A17–A21, A25, B5, B9	Old public archive and accompanying evaluation materials have separate scopes

The original physical-reader result in A20 remains exploratory. A23 adds the common-reader control and a previously examined external cache. A25 separately records 720 new initial states and futures generated after method freezing; it supplies one prespecified field primary and secondary common-output-space diagnostics. The prospective lattice outcomes have their own conditional law and do not independently validate the Brusselator reader. The information witness is not fitted on prospective labels. Early lattice score means are positive despite nonpositive conservative lower bounds; this is no claim of absent gains, an age contrast or the birth of a state. E3 retains its solver-corrected primary-not-confirmed result. Historical numerical checks and proofs below are retained with their original coverage; none was rerun for manuscript integration.

The current evaluation companions are `Evaluation_Replay.zip` (Online Resource 3: `field/` for the two earlier 360-state comparisons and `lattice/` for the frozen information/score endpoints) and `CED_Evaluation_Replay.zip` (the recorded-signal audit in A24). Historical archive roles SA1–SA3 retain their definitions and access scopes in A15, A17 and A18; they are not the numbering of the current Online Resources. The separate anonymized `Prospective720_Evaluation_Replay.zip` (Online Resource 5) accompanies this submission as specified in A25.7; the molecular and raw-trajectory materials retain their stated access boundaries. B9 distinguishes the earlier standalone statistical-package name from the current lattice directory.

Part A. Physical-field diagnostics and supporting evidence

A1. Physical reference and sample construction

For the supplied Brusselator benchmark, the first Neumann spatial mode is cos(x/2). At B = 25/16, its species-space Jacobian is

$$J_{1/2} = \begin{pmatrix} 5/16 & 1 \\ -25/16 & -5 \end{pmatrix}. \tag{1}$$

The critical direction is $(1, -5/16)^T$ and the fast first-mode direction is $(1, -5)^T$. If c_X and c_Y are the corresponding first spatial coefficients, the evaluation coordinates are $a = (16/75)(5c_X + c_Y)$ and $s = c_X - a$. These diagnostics and the supplied cubic amplitude reference are used to interpret a controlled benchmark. They are not labels in the predictive training loss. The known directions do enter the initial ensemble construction, which limits any claim of physics-agnostic discovery.

Modal coefficient and amplitude normalization. The diagnostic coordinates follow directly from a cosine expansion and a two-dimensional change of basis. On the $N = 64$ endpoint grid, write $x_j = 2\pi j/(N-1)$, $\phi_{nj} = \cos(n\pi j/(N-1))$, $w_0 = w_{N-1} = 1/2$, and $w_j = 1$ otherwise. For a sampled deviation field u, the unnormalized DCT-I coefficient and the physical cosine coefficient are

$$C_n = 2\sum_{j=0}^{N-1} w_j \phi_{nj} u_j, \qquad U_n = \frac{C_n}{2h_n}, \qquad h_n = \sum_j w_j \phi_{nj}^2 = \begin{cases} N-1, & n = 0, N-1, \\ (N-1)/2, & 1 \le n \le N-2. \end{cases} \tag{2}$$

Thus $u_j = \sum_{n=0}^{N-1} U_n \phi_{nj}$; the same convention defines V_n from v. In particular, U_0, V_0 multiply the constant spatial mode, whereas U_1, V_1 multiply $\cos(x/2)$, so $c_X = U_1$ and $c_Y = V_1$ above. This fixes the endpoint and species normalization used by the physical readers and error maps in A20 and A25.

Linearization about $(X, Y) = (1, B)$ gives, for wave-number square $q = n^2/4$,

$$M(q, B) = \begin{pmatrix} B - 1 - q & 1 \\ -B & -1 - 16q \end{pmatrix}. \tag{3}$$

At $q = 1/4$ and $B_c = 25/16$, put

$$P = \begin{pmatrix} 1 & 1 \\ -5/16 & -5 \end{pmatrix}, \qquad P^{-1} = \begin{pmatrix} 16/15 & 16/75 \\ -1/15 & -16/75 \end{pmatrix}, \qquad P^{-1} M(1/4, B_c) P = \mathrm{diag}(0, -75/16). \tag{4}$$

Consequently $(U_1, V_1)^T = P(a, s)^T$ yields

$$a = \frac{16}{75}(5U_1 + V_1), \qquad s = -\frac{1}{15}U_1 - \frac{16}{75}V_1 = U_1 - a. \tag{5}$$

The critical left row $\ell_c^T = (16/15, 16/75)$ satisfies $\ell_c^T (1, -5/16)^T = 1$. First-order simple-eigenvalue perturbation therefore gives

$$\left.\frac{d\lambda_c}{dB}\right|_{B_c} = \ell_c^T \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -5/16 \end{pmatrix} = \frac{64}{75}. \tag{6}$$

These projection identities apply to every sampled field. A scalar near-threshold amplitude equation additionally assumes a reduced, slaved-mode description and its amplitude convention. Its supplied cubic coefficient is a separate weakly nonlinear reference, not needed to define a, s, the uniform component or the reported forecast losses. The linear change of basis does not assert that a general simulated field obeys that scalar equation.

The unscaled initial deviations are

$$u_0(x) = (a_0 + s_0)\cos(x/2) + U_X + \sum_{n=2}^{10} c_{X,n}\cos(nx/2),$$
$$v_0(x) = (-5a_0/16 - 5s_0)\cos(x/2) + U_Y + \sum_{n=2}^{10} c_{Y,n}\cos(nx/2). \tag{7}$$

Here $a_0 \sim N(0, 0.018^2)$, $s_0 \sim N(0, 0.055^2)$, $U_X, U_Y \sim N(0, 0.035^2)$, and $c_{X,n}, c_{Y,n} \sim N(0, 0.035^2/n)$ independently before shrinkage. The two uniform coefficients are explicitly sampled, rather than inferred to arise only through evolution. For each sample put $m_X = \min_j u_0(x_j)$ and $m_Y = \min_j v_0(x_j)$, with $M_Y = \max(B - 0.20, 0.20)$. Set $r_X = 1$ if $m_X \geq -0.80$ and $r_X = 0.80/(-m_X)$ otherwise; set $r_Y = 1$ if $m_Y \geq -M_Y$ and $r_Y = M_Y/(-m_Y)$ otherwise. The same $r = \min(1, r_X, r_Y)$ multiplies both deviation fields; the physical initial fields are $X_0 = 1 + ru_0$ and $Y_0 = B + rv_0$. Thus the stated Gaussian scales describe draws before this possible common shrinkage. The known first-mode directions are used across the three controlled parameter regimes.

The KS initial fields use sine and cosine components through mode 8, with random coefficient scale proportional to $1/\sqrt{n}$ for mode n. The field mean is removed and the initial RMS is set to 0.75. The solver then burns in for 50 time units. F0 samples several snapshots of each trajectory at the specified horizons; formal F1b/F2 use separately generated datasets. Train-only standardization and whole-trajectory splitting prevent direct reuse of a trajectory across train and test, while leaving dependence between snapshots within each split.

Table A1. Sample and split units.

Run	Samples	Train / validation / test samples	Split unit
D0, each regime	1,800	1,080 / 360 / 360	Independent initial state
D1	1,800	1,080 / 360 / 360	Independent initial state
E3, E5	3,600 each	60% / 20% / 20%	Independent initial state
F0	90 trajectories × 40	2,160 / 720 / 720	Trajectory, 54 / 18 / 18
F1b, F2	120 trajectories × 30 each	2,160 / 720 / 720	Trajectory, 72 / 24 / 24
F1a, F2a	60 trajectories × 20 each	720 / 240 / 240	Trajectory, 36 / 12 / 12

A2. Prediction networks and target geometry

Table A2. Selected frozen architecture settings. Widths denote fully connected layers; hidden activations are SiLU.

Pipeline	Input / hidden / output structure	Role
D0/D1 representation encoder	128 → 128 → 64 → d	Current field to bottleneck
D0/D1 representation decoder	d → target field	Linear reconstruction or future decoder
F0 representation encoder	64 → 128 → 64 → d	KS current field to bottleneck
F0 representation decoder	d → 64 → 128 → 64	Nonlinear reconstruction or future decoder
F1b sensing predictor	input → 128 → 128 → 64	Separate cheap/full predictors
E5 mask predictor	192 → 256 → 256 → 128	Two-species values plus 64-site mask
F2 mask predictor	128 → 192 → 192 → 64	KS values plus 64-site mask

In D0/D1, AdamW uses learning rate 0.001 and weight decay 10^{-4}, with a 50-epoch cap, batch size 128, and validation patience 8. The three representation model seeds are 20261004, 20261005,

and 20261006. These settings specify a bounded implementation rather than an equal-FLOP comparison to SVD/PCA.

The D1 target for each horizon is centered at the training future mean and divided by a single global training RMS. This choice preserves relative magnitudes across target components. Current input standardization and future target scaling are separate operations. A forecast useful at one horizon need not expose every diagnostic through a linear readout; low critical/fast-coordinate R^2 at the intermediate horizons therefore cannot establish absence of predictive information.

Evaluation ridge probes are fitted on training representations and targets, then applied to the held-out test set. D2 refits the probe in the target regime while freezing the source encoder. The fully frozen predictor follows the source map. Their difference is informative about these two transfer pipelines, but not proof that a learned latent variable has a unique physical meaning.

A3. Conditional scores and measurement budgets

Nearest-neighbor scores use standardized cheap-observation distance to select compatible training examples. In E3, current-state ambiguity is the RMS norm of locally centered standardized noisy full training observations; it is not a noiseless full-state oracle. At the query level, top-scoring states receive full sensing at a fixed fraction of the pool. At the site level, local variance ranks the individual candidate sites, while the regularized linear prediction score measures the penalized-risk reduction derived below.

Table A3. Formal sensing comparisons.

Run	System and horizon	Neighborhood / ridge	Primary budget and contrast
E0	Brusselator	Ensemble prediction dispersion	25% uncertainty versus random upgrades
E3	Solver-corrected Brusselator, τ = 10	k = 24	25% present dispersion versus random upgrades; directionally positive, primary not confirmed (p $\approx$ 0.085)
E4	Brusselator	Local current ambiguity	4 extra sites versus random sites
E5	Brusselator, τ = 10	k = 24, λ = 0.05	Future value, 4 extra sites versus random
E6	Brusselator, target τ = 10	Frozen future-value pipeline	Selection at 10 versus 0.1
F1b	KS, τ = 4	k = 24	25% ambiguity versus random upgrades
F2	KS, τ = 4	k = 32, λ = 0.001	Future value, 4 extra sites versus random

For E5 and F2, the neighbor set supplies the conditioning on the cheap observation; the baseline columns are excluded from the candidate set in the covariance score. E5 includes both species at each selected site. F2 includes one KS value per site. Covariances are computed separately for each query from its training cheap-space neighbors, after local centering with sample denominator $k-1$. In E5, total future variance at or below 10^{-14} returns score zero; the ridge system uses a solve with pseudoinverse fallback. In F2, the denominator is total future variance plus 10^{-12}. These numerical conventions, regularization, neighborhood size and target scaling affect the score.

The mask predictor sees both zero-filled standardized measurements and their masks. Random masks during training expose it to varying counts. E5 uses extra-site counts 0, 2, 4, 8, 16, 32, 62; F2 uses 0, 2, 4, 8, 16, 32, 60. This is measurement-conditioned prediction; it does not establish robustness to all missing-data patterns. For each random repetition, every query independently

receives m distinct sites sampled without replacement from the nonbaseline candidates. The set is not shared across queries. In the Brusselator, both species at a selected site are acquired together. Measurement counts and the prediction pipeline match those of the evaluated policy.

A3.1 Penalized prediction-risk meaning of the score

Within one fixed, centered training neighborhood, let $x = x_S \in \mathbb{R}^r$ denote the selected measurement columns and $f \in \mathbb{R}^m$ the future target. All covariance matrices use the same divisor $k-1$. For a linear reader $W \in \mathbb{R}^{m\times r}$ and ridge coefficient $\lambda > 0$, define the empirical penalized risk

$$R_\lambda(W) = \mathrm{tr}(C_{ff}) - 2\,\mathrm{tr}(WC_{xf}) + \mathrm{tr}\{W(C_{xx} + \lambda I)W^T\}.$$

Writing $A = C_{xx} + \lambda I$ and $W_\lambda = C_{fx}A^{-1}$ gives

$$R_\lambda(W) = \mathrm{tr}(C_{ff}) - \mathrm{tr}(C_{fx}A^{-1}C_{xf}) + \mathrm{tr}\{(W - W_\lambda)A(W - W_\lambda)^T\}.$$

The score numerator is consequently $R_\lambda(0) - \min_W R_\lambda(W)$. It is a decrease in penalized empirical linear prediction risk. The unpenalized empirical MSE at W_λ additionally differs from that minimum by $-\lambda\|W_\lambda\|_F^2$. Nonzero ridge, finite neighborhoods and a separately trained nonlinear evaluation model preclude identifying the score with true conditional variance or realized forecast-error reduction. The normalization/zero-variance conventions above remain exactly those of each implemented experiment. This algebra clarifies the existing score; it changes no selection, prediction or result.

A4. Transfer outcomes and effect definitions

The corrected D2 examples in the main text use d = 1. Below-regime native errors are 0.566369 for the probe and 0.910845 for the direct pipeline. Above-regime native errors are 0.475260 and 0.914266. Compare source-to-target errors with the appropriate native pipeline; direct and refitted-probe predictors are distinct, and a direct-transfer penalty is not universal across directions.

For sensing, the absolute effect is mean random error minus policy error, and the relative effect divides that difference by mean random error. For representation mismatch in F0, the absolute penalty is mismatched error minus matched error, and the relative penalty divides by matched error. These denominators differ deliberately. No pooled cross-system effect is estimated.

E6 preserves the specified primary and reverse secondary comparison. The primary long-target result does not exclude zero in the saved interval. The reverse short-target result is favorable to matched selection. Neither result implies that each horizon owns a unique sensing map. E6 compares paired errors on independent initial-state samples using the saved bootstrap and one-sided within-pair label-swap calculation. Interpreting the latter as a null test requires joint invariance under independently swapping the two labels in each pair, equivalently the corresponding independent sign reversals of the paired differences. Independent initial states alone do not establish this null symmetry.

A5. Calibration, failed pilots, and evidence identities

Run	Identity	Interpretation
E1	Smoke-only negative	Surprise trigger not established in this pilot
E2	Primary unconfirmed plus exploratory discovery; full output not located	Historical OLD SOLVER motivation for E3, not independent confirmation

Run	Identity	Interpretation
Corrected E3	CONFIRMATORY PRIMARY NOT CONFIRMED / DIRECTIONALLY POSITIVE / FROZEN AFTER SOLVER CORRECTION	Directional 1.68% benefit; original 25% primary not statistically confirmed
Initial F1 smoke	Engineering diagnostic	Cheap/full errors approximately 0.926/0.897 leave little headroom
F1a	Calibration only	Selects the longest operable horizon without testing ambiguity allocation
F1b	Formal policy comparison	Independent data/split/noise relative to F1a
Initial F2 smoke	Engineering diagnostic	Full error approximately 1.105 exceeds cheap error approximately 0.953; policy comparison invalid
F2a	Predictor-health check only	Checks measurement-count response without testing a site policy
F2 full	Formal policy comparison	Separate data/split/noise and fixed primary site budget

F1a evaluates τ = 0.25, 1, 2, 4. Its cheap/full error pairs are 0.524170/0.058303, 0.555704/0.104563, 0.616507/0.178057, and 0.739061/0.365567. All meet the recorded operability thresholds, so the longest, $\tau = 4$, is selected. This outcome is not evidence for the ambiguity score.

F2a's errors at extra-site counts 0, 2, 4, 8, 16, 32, 60 are 0.781095, 0.726781, 0.691467, 0.645879, 0.588780, 0.509684, and 0.405072. The saved health gate passes. The formal F2 run retains its own mask-count diagnostics; the two datasets and diagnostic means are not merged.

Primary and secondary comparisons follow the recorded run configurations. These records do not constitute independent verification of a public preregistration date. Software smoke tests and calibration datasets are excluded from formal policy evidence.

A6. Numerical and inferential limitations

The saved F0 step-halving audit includes separately evolved burn-ins, so its horizon-4 mean relative difference, 0.009956, and maximum, 0.058384, include differences in the starting states of the sampled forecasts. F1a retains both its raw and same-initial-state audits. The saved F1b, F2a, and formal F2 mean audit differences are approximately 0.000112, 0.000101, and 0.0000677. They describe different run-specific checks and are not estimates of one pooled solver error.

The historical F0 sign-flip and bootstrap operations resampled snapshots; their nominal values remain legacy diagnostics. A13 supplies a separate exploratory trajectory-based reanalysis of saved later-model errors; it is not a recovered original run or an independent dataset. The allocation p-values answer how favorable the saved policy is relative to equal-budget random allocations on the fixed query pool; they do not provide population uncertainty over new physical trajectories. Secondary budgets, coordinate diagnostics, and selection overlap are not substituted for the primary comparisons.

Retained execution records lack some F-series training logs, complete software-version locks, and checkpoints. The coded-model comparison has separately retained model and parameter provenance, but it does not supply missing historical training records. The F1b validation correlation has no retained supporting source log and is excluded from the reported evidence. F0 result tables are available; its final figure files are unavailable. The supplied material does not claim full execution reproducibility for these empirical runs.

A7. Relation to established methods and scope of comparison

Reference numbers in this section refer to the main manuscript. The table distinguishes established methods from the scope of the tested comparisons.

Component in this study	Established neighboring result	Scope of the present evidence
Predictive bottleneck and horizon comparison	Time-lagged autoencoders and past–future information bottlenecks [1], [2], [15], in addition to [3], [4], (author preprint, arXiv:2312.06608v4)	Selection within the supplied Brusselator ensemble and horizon comparisons in the specified KS data; no information-theoretic optimality claim
Encoder versus evolution-map transfer	Relevant variables and effective dynamics can be identified separately (author preprint, arXiv:2312.06608v4)	The frozen-encoder pipeline refits its target-domain probe; the fully frozen source predictor is a distinct pipeline
Query-pool full-state upgrades	History-based queries for predictive safety and a conditional-entropy/switching-cost objective (author preprint, arXiv:2603.23450v1)	Fixed measurement-count allocation over saved query states; no sequential monitoring or safety guarantee
Conditional future-covariance site selection	Goal-oriented and prediction-oriented sensor design [9], [10], [11]	A training-neighbor ridge proxy evaluated with the frozen mask-aware predictor; no comparison with the cited OED implementations
Greedy optimization of the site proxy	Guarantees depend on the objective; covariance-based sensor criteria can lack submodularity [6], [14]	No general approximation factor is asserted or numerically inferred

The measurement decisions require historical training observations, states, and, for the spatial proxy, training futures. Query futures and realized oracle gains are reserved for evaluation. Measurement counts constrain the sensing budget; feature dimensions constrain the original representation comparisons. A8 separately specifies fixed-width query words for the coded comparisons and capacity theorem. Neither constraint establishes an equal bit rate, physical acquisition cost, or computation cost across the neighboring methods.

The mean-error comparisons with the local heuristic remain descriptive where no dedicated direct test is saved. E6's primary long-target transfer contrast remains unconfirmed. No absence of a matched literature phrase or benchmark combination is treated as evidence of priority. These scope restrictions leave all saved result identities and numerical values intact.

A7.1. Corrected Brusselator sensing evidence

On its saved evaluation pool, the KS primary at the 25% allocation budget improves mean relative L2 error by 4.04% over random allocation; its 5,000 random allocations give one-sided randomization $p = 1/5001 = 0.00019996$. These are the retained primary results, not a new test. Corrected E3 is unconfirmed, and its follow-ups do not support stable positive allocation performance. Independent physical-data repetition of KS F1b is unavailable. The corrected Brusselator E3 primary comparison retains an approximately 1.68% directional improvement at the 25% budget, with one-sided randomization p approximately 0.085.

Solver correction changed future targets and realized sensing gains while preserving the original test pool, scores, and selected samples. The absolute policy improvement increased by roughly 4.3 times, while random-allocation variability increased by roughly 7.5 times. A common-background/relative-error-denominator explanation is supported or plausible; the retained E3 fu-

ture fields do not permit a full decomposition that would establish this mechanism. Random-allocation error intervals are not effect confidence intervals. The Brusselator representation and sensing results use the corrected future-generation pipeline; the initial-state and arithmetic contracts for the separate finite-task certificates are specified below.

A8. Finite-task true-ODE propositions and exact evidence

A8.1. Two sets, two initial-state contracts

The fixed 257-state set is the existing geometric witness selected for target separation. It inherits the old certificate's nearest-cell physical-initial reconstruction and outward IEEE arithmetic contracts. It is not a random population sample selected independently of the geometric search. The fresh 360-state set was generated from the same controlled initial distribution with seed 202610110801 after freezing d4/b4 and all six models; the archived plan and freshness checks establish disjointness from the original 1,800 states. Selection of d4/b4 had already used the prior dimension/bit grid. No model, probe, quantizer, or initial-state distribution is reselected using these 360 outcomes.

For the fresh set, the physical initial point is the exact IEEE binary64 X0/Y0 record. The stored Hermite curve agrees bitwise at the initial endpoint, so the true initial discrepancy is zero; the fixed integer checker uses the uniform positive component upper 2^{-80}, enclosing zero. This does not perturb the initial states or add them as neural inputs. The model input remains the saved float32 sensed-current row. The target is the designated 64-node Neumann/DCT ODE at B = 25/16 and T = 10, rather than the continuous reaction–diffusion PDE. The time-record step is 0.0025. Full trajectory payloads are outside Supplementary Evidence Archive SA1.

The fresh 360 certificates cover 4,000 full intervals per state: 1,440,000 fixed-96-bit propagation checks. The independently evaluated physical-formula segments are fixed at [3,4,5,500,1000,2000, 3000,3996], giving 2,880 audits at 110 decimal digits. Endpoint and β/gap ratios are independently replayed for all 360 states. Maximum β is below 0.384291%; the inherited 257-state maximum is below 0.278222%. All points remain in the risk sums, including negative comparisons and small-norm futures. These β bounds quantify uncertainty around the saved target; they are not predictor errors or sampling confidence levels.

A8.2. Proposition 1: exact finite-task query capacity

Under the inherited 257-state initial/IEEE contract, a common deterministic reader receives only one fixed-width query word, with no state-dependent side information. Common rules and compute are unrestricted. To achieve true-ODE relative L2 $\leq 1/20$ at every one of the 257 states, at least nine effective query bits are necessary. Conversely, the saved current-row matching and future table give a nine-bit oracle with maximum true risk <0.279%. Thus nine bits are necessary and sufficient for this finite task at the stated accuracy.

Proof. The 32,896 certified pair separations exclude one fixed reader output satisfying the 5% contract at two different states. With B ≤ 8 the word alphabet has at most 256 elements, and the pigeonhole principle forces such a collision among 257 states. For sufficiency, unique saved current rows map to 257 different nine-bit words and the reader returns the saved future associated with the word. The target tube bounds the oracle's actual relative error below 0.279% for every state. The two shared raw tables total 263,168 bytes; program, metadata, compute, and workspace are additional. This is an oracle construction, not a learned baseline or a persistent-memory bound. A real-valued scalar could encode 257 labels; no lower bound on real latent dimension follows.

The full construction, quantifiers, separation transfer, and uniform-task corollaries are given in Supplementary Proofs SP1. The exact capacity and oracle endpoints are supplied in SA1 under capacity/. The accuracy-threshold corollary uses the same fixed set and is not a new primary experiment or an E3 parameter-range result.

A8.3. Proposition 2: frozen PRED/RECON true mean ordering

For each of the two fixed sets, define the uniform mean of the paired three-model-seed loss differences, using the true target Y_i and η = 7737125245533627/77371252455336267181195264 (the exact stored binary64 value of 10^{-10}). Conditional on the archived code and reader identities, this mean RECON−PRED is strictly positive. Each seed's paired mean is also strictly positive. The proposition holds separately for saved binary64 forecast vectors and the exact rational readers of those codes with the frozen parameters.

Certified task	PRED mean relative L2	RECON mean relative L2	Mean RECON−PRED	PRED / RECON / unresolved points
Fixed 257-state geometric witness	[0.209014, 0.210791]	[0.881085, 0.883772]	[0.670678, 0.674376]	253 / 4 / 0
Fresh 360 initial states	[0.201611, 0.203304]	[0.880296, 0.882868]	[0.677351, 0.680901]	354 / 6 / 0

Table A4a. Outward deterministic risk enclosures, not confidence intervals. The six decimals shown enclose both readout identities.

Task	Model seed	Certified mean RECON−PRED	PRED / RECON / unresolved points
Fixed 257-state geometric witness	20261004	[0.660191, 0.663870]	244 / 13 / 0
Fixed 257-state geometric witness	20261005	[0.714230, 0.718028]	249 / 8 / 0
Fixed 257-state geometric witness	20261006	[0.637595, 0.641248]	248 / 8 / 1
Fresh 360 initial states	20261004	[0.679436, 0.683028]	346 / 13 / 1
Fresh 360 initial states	20261005	[0.698641, 0.702256]	352 / 6 / 2
Fresh 360 initial states	20261006	[0.653944, 0.657450]	343 / 15 / 2

Table A4b. Outward per-seed difference intervals and pointwise counts. Zero unresolved points in the three-seed average does not imply zero unresolved points in every seed.

Proof outline. Let y_i be the saved numerical target and certify $||Y_i - y_i||_2 \leq \delta_i$. Exact rational square-root enclosures give lower/upper norms for y_i and for each archived forecast-minus-target vector. Forward and reverse triangle inequalities, plus the exact reader-to-saved-output drift where applicable, bound each true numerator. The common denominator $||Y_i||_2+\eta$ lies in [max(η,L_i+η−δ_i), U_i+η+δ_i], where [L_i,U_i] bounds $||y_i||_2$. For the paired numerator difference, use the RECON lower minus PRED upper and RECON upper minus PRED lower. Divide by the positive common-denominator interval with sign-aware endpoints, equivalently its four corners; retain negative lower bounds. Average all point bounds and all paired seed loss bounds. The exact rational mean lower bounds are positive in both identities and for every seed, proving the stated ordering without a bootstrap, p-value, or floating replay tolerance.

Supplementary Proofs SP0 and SP2 give the complete target-enclosure and mean-ordering arguments for the 257-state contract. SP3 applies the same risk inequalities to the separate original-Float64 contract of the fresh 360 states. In the 257 case, the paired mean gap strictly exceeds 335339/500000 and is strictly below 84297/125000. In the fresh case, the outward gap is [0.677351,0.680901]; its exact rational lower is positive. SA1 supplies the exact per-point arrays and certificate endpoints for both identities.

The 257 reverse source indices are [212,809,1012,1376]. Fresh reverse local IDs are [69,86,256, 266,293,319]. The sets and all failures were retained; the 257 and 360 results are not pooled into a 617-state independent experiment. Three seeds are three fixed model pairs, not three independent physical realizations per point. Their average is a loss average, not a 16-bit ensemble forecast.

The 16 effective query bits describe four components quantized at four bits each; the recorded inference parameter payload is 106,416 bytes/model. These models are not the oracle in Proposition 1, and neither their mean error nor the query-capacity theorem proves 5% all-state accuracy, optimal dimension, persistent storage only 16 bits, equal whole-machine compute, or universal PRED superiority. The certificate does not formally verify the whole neural encoder or all hardware-dependent code identities. Fresh independent initial states do not turn a finite deterministic bound into a population risk theorem; KS snapshots in the other experiments remain physically correlated observations, and no independent-Gaussian persistent-memory theorem is invoked here.

A8.4. Supporting attachments and verification scope

Supporting material	Supplied attachment
Physical-target and arithmetic contract	Supplementary Proofs SP0
Nine-query-bit capacity and oracle construction	Supplementary Proofs SP1; SA1 capacity/
Fixed 257-state mean ordering	Supplementary Proofs SP2; SA1 fixed257/
Fresh 360-state mean ordering	Supplementary Proofs SP3; SA1 fresh360/
Exact auxiliary energy matrices and rates	SA1 physical_contract/
Selected empirical result tables	SA1 empirical/
Member identity and rational summary verification	SA1 manifest.json and verify_evidence.py

Supplementary Proofs contain the complete analytic implications from the declared interval enclosures to the stated propositions. Future-Conditioned Representation and Active Sensing in Dynamical Systems (SA1) supplies small certified evidence and a summary checker. The checker verifies supplied hashes, rational endpoints, point-to-mean summation, and target-bound consistency. Full trajectory payloads and complete historical training environments are not included, so it is not a from-zero reproduction of the numerical study. Program-specific paths and administrative metadata are excluded from these reader-facing derivatives, while scientific numeric fields are retained.

The numerical comparison against saved futures and its bootstrap interval remain distinct from these true-ODE deterministic bounds. Corrected E3 retains a directional benefit but its prespecified 25% primary comparison is not statistically confirmed (approximately 1.68%, one-sided randomization p approximately 0.085). The D0 propositions concern internal representation; they do not upgrade E3, establish the old-solver mechanism, or constitute two-system confirmation. Independent-block and range follow-ups retain their exploratory interpretation.

A9. Added cached-field linear and modal analyses

These additional exploratory fits reuse the corrected D0/D1 float32 field caches and original 1, 080/360/360 train/validation/test indices. No trajectories, initial states, or neural training are added. Standardizers, rank construction, and all readout/modal coefficients use training states only; the original held-out test states supply the reported errors. Validation states are unused. The saved neural summary rows remain unchanged. All regimes, dimensions, feature sets, and D1 horizons evaluated here are supplied in SA1's empirical/added_cached_analysis/ tables.

Let X_s, Y_s denote the training-scaled input and future matrices. D0 uses featurewise training means and standard deviations for both, with scale floor 10^{-6}; D1 uses featurewise input scaling and a centered global future RMS at each horizon. The featurewise means and scales and the resulting scaled matrices are float32 for both D0 inputs and targets and for D1 inputs. For D1 targets, the cached future fields are first converted to float64, and the training future mean, global RMS and scaled target are computed in float64. The first fit then uses float64 matrices and centers them again using training means to obtain X_c, Y_c. With $alpha = 0.001$,

$$C = (X_c^T X_c + \alpha I)^{-1} X_c^T Y_c, \qquad X_c C = U \Sigma V^T, \qquad A_d = C V_d.$$

Here V_d contains the first d right singular vectors; d is prescribed without an extra numerical-rank cutoff. For any new physical input x, apply the frozen input scaler to obtain x_s and form $k = x_s A_d$. Using only training codes, set $m_K = \text{mean}(K_{\text{train}})$, $s_K = \max(\text{std}(K_{\text{train}}), 10^{-8})$ componentwise, and $z = (k - m_K)/s_K$. Thus no test example enters the scaler, SVD or code normalization.

For the second fit, let Z be these float64 training codes and $H = [Z\ \mathbf{1}]$. The target F is the raw physical future field, not the scaled first-stage target. The frozen reader is

$$W = (H^T H + \text{diag}(\alpha I_d, 0))^{-1} H^T F, \qquad \hat{f}(x) = [z\ 1]W.$$

Both fits use the unnormalized sum of squared errors and the same fixed numerical penalty 0.001; the reader intercept is unpenalized. Code rescaling changes the effective penalty geometry. This two-stage construction is not asserted to be a standard single rank-constrained ridge optimum or to minimize the held-out mean-relative-L2 loss.

Wehmeyer–Noé's fully whitened linear TAE derivation is TCCA (author preprint pp. 2–3, equations 3–7); its whitening is not the diagonal scaling used here. Chen–Sidky–Ferguson further distinguish variance-weighted nonlinear prediction from slow-mode optimization. No new TCCA/TICA/SRV performance comparison was run, and no exact equivalence to the present normalization is claimed.

For the quadratic diagnostic, six amplitudes are $[a_0, s_0, u_X, u_Y, c_{X,2}, c_{Y,2}]$. Uniform and second-mode coefficients use endpoint-half trapezoidal cosine projections on the fixed 64-node grid; a_0 and s_0 are computed with the benchmark's first-mode species directions. Fit future critical amplitude using (i) a_0, (ii) all six linear features, (iii) those features plus the 21 degree-two monomials with $i \leq j$, and (iv) additionally the 56 degree-three monomials with $i \leq j \leq k$. Each ordinary least-squares fit has an intercept. Test R^2 is 0.977506, 0.977350, 0.999547, and 0.999786, respectively. Physical-coordinate information is available to this diagnostic; it is not a label-free learned baseline or a neural-mechanism certificate.

For D1, the actual multihorizon cache differs from the D0 data used by the downloaded diagnostic. Project the future field into critical first-mode, fast first-mode, uniform, and residual components; divide each component's squared centered training norm by the total squared centered future norm. Critical/fast species directions are nonorthogonal, and the unweighted norm is not the trapezoidal cosine inner product, so these ratios do not sum to an additive variance decomposition. At τ = 0.5 and 2, uniform ratios are 0.770716 and 0.912596. Predicting the scalar d = 1 RTRP code from both current uniform species amplitudes yields held-out R^2 = 0.999692 and 0.999954. The direction of this regression is uniform amplitudes to linear code. The original neural latent arrays/checkpoints are absent; it supplies no direct PRED uniform-coordinate R^2.

The input hashes, source-script hashes, exact numerical tables, and cached analysis implementation are archived separately. Deterministic PCA/oracle/full-linear control replays across all three regimes differ from the saved summary by at most 3.51×10^{-9} in future relative L2. This agreement checks cache and metric use; it is not independent re-integration or a new neural replication.

A10. Independent-block and short-horizon ambiguity outcomes

Ten regenerated blocks under the original E3 condition give mean relative advantage -0.168554%, an approximate block t95 interval $[-0.990179\%, 0.653071\%]$, and four positive/six negative blocks. The 1.68% corrected single-pool primary remains its own frozen result; it is outside this exploratory interval for the repeated-pipeline mean. No equivalence margin was specified, so this does not establish zero effect.

The separate nine-cell grid reuses ten physical-data blocks across horizon/noise cells, with predictors retrained under each condition. It is 90 fits/conditions, not 90 independent physical blocks. Below, positive means lower ambiguity-policy error than equal-budget random error. Intervals use the saved small-sample t approximation with Bonferroni .05/9 across all nine cells.

Horizon	Noise SD	Mean relative effect (%)	Approximate family95 interval (%)	Positive blocks
0.1	0.01	-4.402	[-7.330, -1.474]	0/10
0.1	0.03	-3.532	[-5.533, -1.530]	0/10
0.1	0.06	-2.892	[-4.096, -1.687]	0/10
2	0.01	-0.386	[-1.514, 0.742]	4/10
2	0.03	0.169	[-1.058, 1.396]	5/10
2	0.06	0.487	[-0.225, 1.200]	8/10
10	0.01	-0.378	[-1.743, 0.987]	3/10
10	0.03	0.295	[-0.846, 1.436]	6/10
10	0.06	-0.470	[-1.818, 0.877]	5/10

Table A5. All nine exploratory conditions, B = 1.625, 25% upgrades, k = 24. All short-horizon adjusted intervals lie below zero for the relative-L2 effect. The absolute-L2 diagnostic need not have the same sign, and all-cell results depend on the fixed training pipeline and small-sample interval approximation. No continuous negative region, universal ineffectiveness, or causal mechanism follows. KS F1b has the favorable saved-pool primary reported in A7.1, without comparable independent-block repetition.

A11. Incremental quadratic-probe interpretation

The six-feature linear probe has test R^2 = 0.9773499911097615 and the quadratic-augmented probe R^2 = 0.999546854021. Both use the same held-out future critical amplitude and the same centered test variance denominator. The quadratic augmentation reduces the residual squared error left by the six-feature linear fit by $(R^2_{\text{quadratic}} - R^2_{\text{linear}})/(1 - R^2_{\text{linear}})$ = approximately 0.979994, or 98.0%. This describes the incremental residual reduction, rather than attributing all already-explained amplitude variation to quadratic terms. It is a saved-data exploratory probe comparison, not a causal identification of the neural encoder's mechanism.

A12. Restored-pipeline POD–QR comparisons

The original E5/F2 data arrays and mask-aware checkpoints were not retained. The added comparison therefore regenerated the original data seeds and retrained the original predictors in a separate run, preserving the two systems' architectures, normalizations, three model seeds, early stopping, and epoch caps. No original results or checkpoint objects were overwritten. E5 uses 3,600 independent initial states, seed89101, split89201, noise89301, corrected step0.0025 and horizon10. F2 uses 120 trajectories $\times$30 snapshots, data98201, trajectory split98301, noise98401, step0.25, burn-in50 and horizon4. Their original 2,160/720/720 splits are preserved. These are restorations of the original single datasets, not fresh-data repetitions of sensing policies.

All original function/class definitions in the adapted execution source match their original ASTs. The unchanged training prefixes terminate after the original three networks are trained, and a new continuation saves arrays/checkpoints and evaluates four-extra-site policies. Prior numerical-audit records are reused; no extra audit integrations, randomization tests, bootstrap, KS independent blocks, or D1 training are added. The actual restored execution uses Python3.11.16, PyTorch2.14.0+cpu, NumPy2.4.6 and SciPy1.17.1. This runtime is declared rather than assumed identical to every historical run.

POD uses centered standardized noisy training-current fields. Its rank equals the final scalar reading budget: r12 for two-species E5 with two baseline plus four extra sites, and r8 for scalar F2 with four baseline plus four extra sites. QR does not use test futures, realized gains, or test scores. Force the baseline columns before choosing new sites. For KS, remove the baseline column

span and apply QR column pivoting to the remaining basis columns, taking four pivots. For E5, two species columns form an inseparable site block: at each step select the remaining block with largest summed squared residual norm, then remove the full selected-column span. Exact score ties choose the smaller site index. This is a group-constrained adaptation of POD–QR, rather than Manohar's scalar algorithm unchanged. The selected zero-based extra sites are [35,29,22,44] in E5 and [54,5, 21,38] in F2.

The primary predictions preserve the original ensemble arithmetic: E5 averages its float32 scaled forecasts before inverse scaling; F2 accumulates scaled forecasts in float64 before inverse scaling and a float32 cast. Error is computed after prediction averaging, not by averaging seed losses. The following are individual-seed diagnostics.

System	Model seed	Greedy mean relative L2	QR mean relative L2	Lower mean
Brusselator	20261004	0.886955	0.929436	Greedy
Brusselator	20261005	0.890861	0.921170	Greedy
Brusselator	20261006	0.877227	0.919155	Greedy
KS	20261004	0.512192	0.535903	Greedy
KS	20261005	0.514316	0.502029	QR
KS	20261006	0.509744	0.505286	QR

Table A6. Fixed individual-seed mean errors, distinct from the forecast-averaged primary comparison. All model seeds and test points remain in the results; no unfavorable seed is removed.

The mean ensemble absolute-L2 errors are 0.129051 (greedy) versus 0.137355 (QR) in E5 and 4.786517 versus 4.852568 in F2. Separate Float64 norms computed from all saved forecasts and numerical future fields give mean QR−greedy relative-error differences 0.0384012195 and 0.0040683791, preserving the original computed signs. This check does not improve physical future-generation accuracy or establish statistical confirmation. Mean errors from reconstructed original predictors differ from the archived greedy means by 5.96×10^{-8} and about 5.55×10^{-17}; best validation scores differ slightly, while the saved best epochs match (E5:6/10/7; F2:37/34/42). Cross-runtime reconstruction is not claimed bitwise identical.

All 720 point differences, per-seed tables, and KS trajectory-mean differences are supplied in Future-Conditioned Representation and Active Sensing in Dynamical Systems (SA2), with the train-only selector implementation and payload hashes. Trajectory means are descriptive, not a cluster confidence interval. The generic legacy checkpoint-replay field name mentions GPU, but this run compares CPU models and CPU reloads; no GPU was used. Original confirmatory labels inside the reconstructed old configs describe the original recipe, while the added QR comparison is exploratory. No population, seed-uniform, or universally optimal sensing conclusion is drawn. E3 and the D0 finite-task certificates retain their previous statuses.

A13. Trajectory-based F0 reanalysis

The original F0 primary has matched mean relative L2 0.5506003202088593 and mismatched mean 0.6988650810392351, with snapshot-bootstrap interval [0.1399013351884579, 0.15669102239885657] and snapshot sign-flip value 0.0001999600079984003. Those inferential values did not preserve within-trajectory dependence and are retained here for historical provenance, not current trajectory-level confirmation.

The original cached fields, split indices, trajectory IDs, and snapshot IDs survive. The original run did not save its per-snapshot primary errors or checkpoints. A subsequent finite-code pilot retrained the same six networks with the original recipe and seeds on that unchanged cache, saving its unquantized reference errors, scalers, latents, readouts, and checkpoints. The current analysis uses only those saved unquantized errors; no new model fitting, training, trajectories, or integration

is performed. Their two means differ from the original primary by less than 5e-10; agreement of means is not proof of pointwise historical identity.

For each of the 720 test snapshots, average the three paired seed losses; for each of 18 trajectories, average its 40 snapshots. Resample 18 complete trajectories with replacement 100,000 times, NumPy default_rng seed 202610120801, and take the 2.5% and 97.5% quantiles with linear interpolation. Models remain fixed. The paired gap is 0.14826476075698022, relative penalty 26.92783774796%, and the 95% percentile interval is [0.13334732778656577, 0.16112307158776026]. All 18 trajectory mean gaps are positive, ranging from 0.059477394708660224 to 0.1823884331318979. The interval assumes independent test trajectories conditional on the fixed models and recorded generation protocol. It has only 18 clusters, excludes training variability, and is exploratory on previously examined data.

As a conditional calculation, the all-positive trajectory gaps give one exceedance in the 2^{18} sign-flip orbit: one-sided p = 1/262144 = 0.000003814697265625. The exact count uses positivity of the exact dyadic rational values of the stored cluster means: the observed all-plus sum is the unique orbit maximum. Interpreting this as a null test additionally requires joint invariance under independently reversing trajectory signs. Pairing or independent trajectories alone does not establish that symmetry. It is not an unconditional exact test or an independent KS confirmation. The bootstrap interval is a sampling calculation, not a deterministic ODE error certificate.

SA3 supplies all 720 paired seed-error rows, the 18 trajectory rows, data mappings, and the standalone analysis script. Its output hashes identify the later-model source. Historic RRR keys in unchanged analysis tables map to the public RTRP label; no numerical result is changed by renaming.

A14. Complete Brusselator time step and diagnostic coverage

The corrected step is $D(h/2)R_{\text{Heun}}(h)D(h/2)$. D is the exact diffusion map in the unorthogonalized DCT-I/inverse-DCT-I pair on endpoint samples; mode n has wave number $n\pi/L$ and multiplier $\exp(-D_{\text{species}}k_n^2 h)$, with species diffusivities 1 and 16. Adjacent half diffusions combine. The reaction substep is explicit Heun. This Strang composition is second order for smooth solutions; it is not the unsymmetrized full-reaction/full-diffusion map.

Saved diagnostics at h=0.0025 versus h/2=0.00125 are:

Diagnostic	Samples / scope	Mean relative field difference	Maximum
D0 audit	32 diagnostic initial states, tau=10	0.0000242649	0.000139089
D1 audit	32 diagnostic initial states, tau=10	0.0000329321	0.000224218
E5 audit	Run-specific diagnostic subset	0.0000252171	0.0000914332

The independent nonlinear solver-validation set contains 18 diagnostic states. Its aggregate reference-error ratios on halving 0.005 to 0.0025 and 0.0025 to 0.00125 are 4.00284687 and 4.00596681 (observed orders 2.00102643 and 2.00215047). At 0.0025 its maximum relative deviation error is 0.000133742356. The spatial comparison uses only three same-initial-state cases. These finite diagnostics are not checks of every training/test state or frozen forecast loss, and are separate from the supporting true-ODE certificates. KS F0's saved audit includes separately integrated burn-ins and is not a same-initial-state target-sensitivity check.

A15. Reproducibility package and naming

SA3 includes the actual snapshots and checkpoints listed in the main availability statement, not only their hashes. Its README contains relative-path commands and a coverage matrix. The runnable

core covers cached linear/modal analyses, saved-prediction losses, QR selection, trajectory reanalysis, and saved-model inference checks. Regeneration/training sources are supplied separately and not executed by the replay commands. Historical D1 neural latents remain absent. Generator availability does not imply recovery of old execution environments or successful reruns of every historical result. The reproducibility archive for *Future-Conditioned Representation and Active Sensing in Dynamical Systems*, version 1 (Zenodo, 2026), is available at `https://doi.org/10.5281/zenodo.23599714`. Supplementary material 3 provides the access information; the deposited ZIP retains its original document mirrors.

The linear comparator is rank-truncated ridge predictive regression (RTRP), specified by its centering, full ridge fit, target-fit SVD truncation, and training-fitted evaluation readout in A9. It is a descriptive name for the actual algorithm; no equivalence to a standard constrained ridge reduced-rank-regression optimum is asserted. Unchanged legacy CSV fields/functions named RRR remain explicit aliases in the release mapping.

A16. Degenerate archived global-variance selector

The archived E4 global_mask and E5 global_variance_mask sum the two species' full-training-set feature variances after training-feature standardization. For every nonconstant column standardized on that same set, the exact marginal variance is one, so the site score is two. On the supplied E5 training arrays, the float32 score ranges only from 1.9999974966 to 2.0000035763. The selected four sites change from [5,21,60,15] to [42,5,41,15] when only accumulation precision is changed. Raw noisy training-field variances instead select [3,62,2,1]. These are training-score diagnostics, not evaluations of a corrected policy's forecast loss.

The E4 archived GLOBAL_VARIANCE_4 output 1.002880, compared with BASE_ONLY 0.971876, is retained solely as a degenerate-implementation record. It is excluded from physical global-variance interpretation; no corrected E4 value is inferred from E5. Original source definitions, CSV fields and nested evidence remain unchanged for provenance, with this annotation applying wherever the legacy GLOBAL_VARIANCE label occurs. The local-neighborhood variance primary and the future-covariance and POD–QR selectors do not have this equal-training-marginal-variance identity. Their saved independent comparisons retain their original contracts. No new global-policy loss, mask-aware model, integration or fit is introduced.

A17. Unquantized D0 recovery-model replay

The critical-parameter D0 d4 replay supplies the three PRED and three RECON checkpoints saved by the later same-recipe recovery on the original 1800-state cache. It is restricted to B = 25/16, the original 1080/360/360 split, and mean relative L2 averaged over the three seed losses. It is neither the original historical weight identity nor new physical-data confirmation, and does not cover the other two B values in the archived three-regime D0 summary (Supplement A9; SA1, `empirical/brusselator/D0/representation_summary.csv`).

The release includes the input scalers and indices, float train/validation/test latents, latent scalers and frozen future-field ridge readers. No fitting is invoked. The architecture class is an unchanged extraction of the supplied original class definition. Run python scripts/replay_d0_float.py –root . –outdir replay_D0_float in a new output directory. It verifies data splits and checkpoint metadata, computes CPU encoder outputs, checks them against the saved float latents, applies the frozen normalization and reader, and retains all six models' forecasts and point errors. Saved-latent reference losses are separately checked against the recovery CSV.

The saved PRED reference mean is 0.03877916558099172, compared with the historical summary mean 0.03877916567013193; CPU input-to-forecast replay gives approximately 0.03877916348. The saved and CPU RECON means are approximately 0.86284047783 and 0.86284046813. These small runtime differences do not establish historical weight identity. Full

per-seed values, current runtime and declared numerical comparison tolerances are retained with the replay output; those tolerances are not confidence intervals or true-ODE error bounds.

A18. Numerical-certification implementation and payload contract

The evidence archive's certification directory supplies the retained interval, Bernstein-residual, propagation, integer-check and high-precision physical-audit source snapshots together with the fixed257 and fresh360 execution adapters. A source manifest identifies the original and released program bytes; machine-specific path strings are projected to a documented relative layout, without changing numerical constants or mathematical kernels.

The accompanying payload contract distinguishes supplied small certificate endpoints from the large per-state trajectories and interval-segment payloads needed to reconstruct a full physical audit. The source snapshots document that computation; the portable summary check still performs no integration or millions-of-segment audit. This materials release does not claim that all historical execution environments or every full physical payload has been reproduced from scratch.

A19. Data, solver, model and result crosswalk

The corrected D0/D1/E3 run records date to 7 October 2026. Data hashes, source hashes, configurations and completion logs link the following pipelines; cache identity alone is not an ODE accuracy certificate. The D0/D1 caches used by the later modal analyses and the archive have matching SHA256 identities.

Table A7. Pipeline identities and inferential scope.

Result	Physical data / target	Model or analysis	Scope
D0; original field-study representation tables	Three solver-corrected 1, 800-state caches; 64 physical endpoint nodes, DCT-I, Strang–Heun, step 0.0025, horizon 10; original splits.	Corrected-run PRED/RECON summaries; later RTRP and full-linear fits use the same training cache.	Empirical unquantized comparison; not an all-linear-method lower bound. Historical checkpoint identity is incomplete.
D1; original field-study horizon table	Independent corrected 1, 800-state multihorizon cache, horizons 0.1/0.5/2/10; same corrected step and original splits.	Original corrected-data neural summaries; later field, uniform-mode and RTRP diagnostics use this cache.	No retained historical neural latents; field/linear diagnostics do not identify the neural coordinate.
Scalar modal probes	Corrected critical D0 cache; current field to future critical amplitude.	Training-only least squares with six linear features and 21 quadratic/56 cubic products; held-out evaluation.	Scalar predictability diagnostic; 98.0% concerns the six-linear residual, not the neural mechanism or full-field error.
Corrected E3 primary	Original physical seeds, split and noise specification; regenerated futures from solver_v2 at step 0.0025.	Completed logs record cheap/full training for all three seeds. Test pool, ambiguity scores and selected 180 states retain their verified identity.	Not just rescoring frozen old predictors. One-sided p=0.084983: primary not confirmed; follow-ups remain separate.
Later D0 recovery models	Same corrected critical cache and original 1,080/360/360 split.	Later same-recipe three PRED/three RECON models, scalers, float codes and frozen readers; lineage in configs/D0_recovery_lineage.json.	Supports the saved replay pipeline; does not recover historical weight identity or provide fresh-data confirmation.
257/fresh360 certificates	True futures of the specified 64-node Neumann/DCT ODE; distinct initial-state and arithmetic contracts in SP0–SP4.	Frozen d4/b4 (16-query-bit) models, three paired seeds and two reader identities on fixed finite sets.	Deterministic bounds for coded pipelines; no strict certification of the unquantized tables, scalar probes or sensing results.

A20. Same-task modal and neural field comparison

A20.1 Data identity and frozen choices

This exploratory diagnostic reuses the corrected critical-regime D0 cache at $B = 25/16$ and horizon $\tau = 10$. Its 1800 independent initial states retain the original 1080/360/360 training/validation/test split. Validation states are unused. The 360 evaluation states are the original test split, distinct from the fresh 360-state true-ODE certification set. Features, ridge coefficient, ordinary-least-squares objectives and the three saved PRED seeds were fixed before this diagnostic was fitted. There is no new physical integration, sampling, neural training or neural-network inference. The supplied replay applies the stored coefficients of the three readout fits without refitting.

Every method predicts the same raw 128-component future deviation field. Metrics promote the saved targets and predictions to float64 and use the denominator $d_i = \|y_i\|_2 + 10^{-10}$ for test state i. The full-state linear positive control gives mean relative L2 0.1520875391 under the historical native-float32 denominator, agreeing with the retained original value within 1.2×10^{-16}; under the common float64 denominator it gives 0.1520875386. This numerical convention does not change its six-decimal table entry.

A20.2 Physical feature maps and fitted readers

Let $x_j = 2\pi j/63$ for $j = 0, \dots, 63$, with trapezoidal weights $w_0 = w_{63} = 1/2$ and all other $w_j = 1$. Define $\phi_{nj} = \cos(n x_j/2)$ and, for the current deviation fields $u = X - 1$ and $v = Y - B$, the coefficients

$$U_n = \frac{\sum_j w_j \phi_{nj} u_j}{\sum_j w_j \phi_{nj}^2}, \qquad V_n = \frac{\sum_j w_j \phi_{nj} v_j}{\sum_j w_j \phi_{nj}^2}, \qquad n = 0, 1, 2. \tag{8}$$

The first-mode amplitudes a and s satisfy $(U_1, V_1) = a(1, -5/16) + s(1, -5)$. Hence the six-feature map is

$$a = \frac{16}{75}(5U_1 + V_1), \quad s = U_1 - a, \quad f = (a, s, U_0, V_0, U_2, V_2), \quad F_{27} = (f, (f_p f_q)_{1 \le p \le q \le 6}). \tag{9}$$

For either six or 27 features, a column of ones is appended and unregularized float64 least squares maps the training feature matrix to all 128 future components. No feature selection, cross-validation or test-based tuning is used. The design matrices have full ranks 7 and 28, with condition numbers approximately 56.3 and 2326. The full-state linear reader instead retains the two-stage float32 scaling, ridge coefficient 0.001 and unpenalized intercept of the original reference. All scalers are fitted on training states. The saved coefficients specify the precise arithmetic replay.

The PRED comparator is the later same-recipe unquantized recovery pipeline of A17, evaluated from its three already saved forecasts. No neural forward pass is needed for this diagnostic. Means average losses, rather than forming an ensemble forecast. Per-seed field errors are 0.040395, 0.036492 and 0.039450 for seeds 20261004, 20261005 and 20261006, respectively. Their mean is 0.038779, compared with 0.022884 for the quadratic reader. Comparing the quadratic loss with the mean of the three PRED losses at each state gives 288 quadratic-favorable and 72 PRED-favorable cases, with no ties. These counts have no assigned p-value or confidence interval.

A20.3 Resource accounting and target-dependent accuracy

The six-feature and quadratic readers contain $(6 + 1)128 = 896$ and $(27 + 1)128 = 3584$ fitted coefficients. The full-state linear reader contains $(128 + 1)128 = 16512$. Each PRED future reader contains $(4 + 1)128 = 640$ coefficients, in addition to 25028 encoder weights and biases in the $128 \to 128 \to 64 \to 4$ network. Its original decoder is unused in this probe comparison. Analytic modes carry physical prior information and the quadratic map is a fixed nonlinear expansion. These coefficient counts do not equate real latent dimension, query bits, shared memory, runtime or

training cost. In particular, the 27-feature result is not a four-coordinate competitor with matched resources. The base observations are six current amplitudes, not 27 independent sensed quantities: F_{27} is their deterministic feature expansion. For any finite set, a centered affine output matrix factors as $Z_c B$ and has rank at most the number of feature columns. Thus the PRED reader permits centered output rank at most four, the six-mode reader at most six, and the expanded reader at most 27. These are algebraic upper bounds, not measured ranks. Feature expansion changes the admissible output span as well as the coefficient budget, so this comparison cannot separate physical prior, nonlinear dictionary and output-rank effects under equal resources.

main Table 2 reports projections of each full-field forecast onto the critical and uniform components. The quadratic map predicts the full field more accurately but its critical-amplitude R^2 (0.999547) is below the mean PRED value (0.999673). The joint-uniform R^2 instead favors the quadratic map (0.999806 versus 0.996637). A fixed linear output projection commutes with ordinary least squares, so the full-field modal readers reproduce the earlier scalar-probe R^2 values 0.977350 and 0.999547 without an additional scalar fit. The older cubic scalar probe remains a different diagnostic.

A20.4 Error components and cross terms

For each field error $e_i = \hat{y}_i - y_i$, the same coefficient maps define its critical component e_i^c, fast component e_i^s and uniform component e_i^u. The remaining component is $e_i^r = e_i - e_i^c - e_i^s - e_i^u$. Critical and fast components use species directions $(1, -5/16)$ and $(1, -5)$ with spatial factor ϕ_1; the uniform component repeats U_0, V_0 at every site. These maps use trapezoidal coefficients, while the loss uses the unweighted spatial Euclidean norm. The decomposition is not orthogonal. With $J = \{c, s, u, r\}$, the exact bookkeeping identity is

$$\frac{1}{n}\sum_i \frac{\|e_i\|_2^2}{d_i^2} = \sum_{j\in J}\frac{1}{n}\sum_i \frac{\|e_i^j\|_2^2}{d_i^2} + 2\sum_{j<k}\frac{1}{n}\sum_i \frac{\langle e_i^j, e_i^k\rangle}{d_i^2}. \tag{10}$$

All six cross terms are retained in the supplied component CSV. This identity concerns mean squared relative error, not the main mean relative L2, nor the square of its mean. For the quadratic map and PRED seed average, respectively, normalized critical squared-norm contributions are 0.0012103 and 0.0006969; uniform contributions are 0.0000546 and 0.0009906; remaining-field contributions are 0.0000233 and 0.0007940. Their total squared relative errors, including cross terms, are 0.0012888 and 0.0025168. Thus the component diagnostics support a target-dependent accuracy distinction, not a percentage partition of the mean-L2 advantage.

A20.5 Verification and inferential limits

An independent implementation used DCT-I coefficients in place of explicit weighted cosine sums, and pivoted/augmented QR in place of the retained least-squares/ridge solves for numerical verification of the same fixed objectives. Maximum prediction discrepancy was below 4×10^{-16}. Saved-coefficient replay reproduces the stored predictions; component and cross-term means close within 7×10^{-18}. The release includes readouts, predictions, per-state losses, the metric and feature definitions, fit-source snapshots, and a no-fit replay command. This is a numerical consistency check of the saved solver-future diagnostic, not strict certification of the three new readers.

The original test set has already been inspected in earlier analyses. No new inferential test, new-data replication, causal mechanism identification, optimality statement or minimal-state claim follows from the added comparison. The 257-state and fresh360 certificates in Supplementary Proofs SP0–SP4 concern different frozen coded neural pipelines and true-ODE contracts; they do not certify the new modal readers or the unquantized table. The corrected E3 status and all sensing results are unaffected.

A21. Molecular representation comparison

A21.1 Data, targets and separation of trajectories

The public mdshare alanine-dipeptide set contains three independent 250 ns trajectories, each with 250000 frames separated by 1 ps, and ten heavy atoms. The source metadata describe simulations at 300 K with AMBER ff99SB-ILDN, TIP3P water and a 2 fs Langevin integration step. We analyze the supplied data, without generating new molecular dynamics. The three source files total 96002901 bytes. Their upstream URLs, SHA256 hashes, CC BY 4.0 attribution and topology are listed in `molecular/source_manifest.json` in the original field-study repository archive cited below. The molecular TAE/TICA/PCA precedent is Ref. [1], whose author version pp. 5–6 describes the three trajectories and aligned heavy-atom input; this is not a first molecular application of predictive encoding.

Each frame is centered and aligned by a proper-rotation Kabsch transform to the first frame of that fold's training trajectory. The input is the resulting 30 Cartesian components. Coordinates are in nm; topology coordinates are in angstroms. The supplied dihedrals were independently recomputed from the coordinates and topology before fitting. All input and target means and isotropic scalar scales use training frames only. The future coordinate target F has 30 components. The future torsion target T is $(\cos\phi, \sin\phi, \cos\psi, \sin\psi)$, where ϕ and ψ are the two backbone dihedrals. PRED_T uses these angle labels as its explicit training target; PRED_F does not. Both physical feature maps below are computed from the same current coordinate observation using known topology.

Trajectory indices are zero based. Training/validation/test roles are $(0, 1, 2)$, $(1, 2, 0)$ and $(2, 0, 1)$. Training and validation use starting indices $0, 5, \ldots, 249695$ (49940 pairs each); testing uses every starting index from 0 through 249699 (249700 pairs per role rotation). The common last start leaves room for the largest lag, 300 ps. No trajectory is split between roles within a fold. Across folds, trajectories exchange roles: the folds are not three independent studies, and their 749100 test starts are not independent physical realizations.

A21.2 Representation fits and shared evaluation readers

The fixed dimensions are $d = 2, 4$ and lags are 3, 30 and 300 ps, with primary $d = 2$, lag 30 ps. Each neural objective uses seeds 20261004, 20261005 and 20261006. The encoder is $30 \to 128 \to 64 \to d$, with SiLU hidden activations and a linear latent layer. A linear training head maps the code to the present coordinates (RECON), future coordinates (PRED_F) or future torsion components (PRED_T). AdamW uses learning rate 0.001, weight decay 0.0001, batch size 512, at most 50 epochs, early-stopping patience 8, and minimum validation improvement 10^{-8}. The fixed grid contains 126 neural fits: 18 RECON encoders (three folds, two dimensions and three seeds), each reused across all three forecast lags, plus 54 PRED_F and 54 PRED_T encoders fitted separately for each lag. The RECON objective has no future-lag dependence. All encoders, baseline projections and readout hyperparameters are fixed before any test loss is evaluated. The retained execution audit verifies that ordering. This section reports those completed fits; manuscript integration introduces no further fitting or model inference.

Each representation has an affine ridge readout for F and another for T. Latents are standardized on the training pairs, with the intercept unpenalized. The objective is mean squared Euclidean error plus $\alpha\|W\|_F^2$, with α selected from $10^{-6}, 10^{-3}, 1$ by validation error; a tie selects the larger value. Networks use float32; readouts and evaluation use float64. For each target, NMSE is the mean squared Euclidean test error divided by the total variance of its training future target. It is not the mean of per-state relative L2 errors used for the fields. Normalization differs between F and T, so their numerical levels are not directly comparable.

PCA uses the training-coordinate covariance. Reversible TICA uses current/future coordinate pairs with the same training-input centering and scalar scale. It symmetrizes their second moments and lagged cross moments, adds $10^{-6}\,\mathrm{tr}(C_{00})/30$ to the diagonal of C_{00}, and selects the largest algebraic generalized eigenvalues. RTRP_F/T first fits a full coordinate-to-target ridge regression and

then truncates its fitted training targets by singular-value decomposition to the specified rank. The retained projected coordinates receive the same evaluation readouts as other representations. This is the actual rank-truncated ridge predictive regression comparator, not a claim of equivalence to a standard constrained ridge-RRR optimum. FULL_LINEAR uses all 30 aligned input components. TRAIN_MEAN and PERSISTENCE use the training mean and current observation, respectively, without a fitted latent.

P4 is $(\cos\phi, \sin\phi, \cos\psi, \sin\psi)$. P12 appends cosine and sine at each argument $2\phi, 2\psi, \phi+\psi, \phi-\psi$. It has twelve real features, not two. The neural encoders contain 12354 parameters at $d = 2$ and 12484 at $d = 4$, excluding their training and evaluation heads. A coordinate head has 90 or 150 parameters; a torsion head has 12 or 20. P4/P12 and FULL_LINEAR therefore provide distinct feature/prior budgets, not equal-resource neural competitors. No quantization, persistent-bit bound or FLOP matching is imposed.

A21.3 Complete mean-error matrix

The following table retains all dimensions and lags, including larger-budget references. Neural entries first average the three seed losses within a fold, then give equal weight to the three role rotations. Deterministic references average the fold losses directly. The molecular CSV contains the individual seed/fold values and current-angle diagnostics. No ensemble averaging of predictions is used.

Table A8: Molecular mean errors for all saved method/dimension/lag combinations.

Method	Features	Lag (ps)	F NMSE	T NMSE
TRAIN_MEAN	0	3	1.002201	1.002838
PCA	2	3	0.440427	0.428323
PRED_F	2	3	0.372871	0.357456
PRED_T	2	3	0.373765	0.356403
RECON	2	3	0.440678	0.428524
RTRP_F	2	3	0.386737	0.371991
RTRP_T	2	3	0.386905	0.371662
TICA	2	3	0.387236	0.371943
P4	4	3	0.375242	0.355676
PCA	4	3	0.374466	0.357791
PRED_F	4	3	0.355212	0.337837
PRED_T	4	3	0.356328	0.337362
RECON	4	3	0.374466	0.357793
RTRP_F	4	3	0.371141	0.354164
RTRP_T	4	3	0.371413	0.353941
TICA	4	3	0.371279	0.354303
P12	12	3	0.359779	0.341467
FULL_LINEAR	30	3	0.371078	0.353932
PERSISTENCE	30	3	0.574533	0.528993
TRAIN_MEAN	0	30	1.002800	1.003119
PCA	2	30	0.746456	0.740097
PRED_F	2	30	0.671529	0.658285
PRED_T	2	30	0.671752	0.658095
RECON	2	30	0.746756	0.740409
RTRP_F	2	30	0.682229	0.669650
RTRP_T	2	30	0.682274	0.669616
TICA	2	30	0.682196	0.669581
P4	4	30	0.684573	0.671662
PCA	4	30	0.684298	0.671960
PRED_F	4	30	0.671234	0.658073
PRED_T	4	30	0.671248	0.658118

Method	Features	Lag (ps)	F NMSE	T NMSE
RECON	4	30	0.684299	0.671961
RTRP_F	4	30	0.682216	0.669667
RTRP_T	4	30	0.682268	0.669689
TICA	4	30	0.682096	0.669550
P12	12	30	0.672712	0.659509
FULL_LINEAR	30	30	0.682228	0.669685
PERSISTENCE	30	30	1.062477	1.033982
TRAIN_MEAN	0	300	1.002859	1.003178
PCA	2	300	1.000244	1.000166
PRED_F	2	300	0.946831	0.939247
PRED_T	2	300	0.947034	0.939475
RECON	2	300	1.000679	1.000650
RTRP_F	2	300	0.952510	0.945724
RTRP_T	2	300	0.952495	0.945708
TICA	2	300	0.952429	0.945648
P4	4	300	0.954434	0.947976
PCA	4	300	0.953314	0.946672
PRED_F	4	300	0.946960	0.939384
PRED_T	4	300	0.946998	0.939429
RECON	4	300	0.953311	0.946671
RTRP_F	4	300	0.952587	0.945786
RTRP_T	4	300	0.952577	0.945791
TICA	4	300	0.952403	0.945617
P12	12	300	0.947117	0.939559
FULL_LINEAR	30	300	0.952591	0.945780
PERSISTENCE	30	300	1.847044	1.827302

At the primary cell, own-target neural errors are only slightly below those of P12, with different budgets and no superiority or equivalence test. At the shortest lag and $d = 2$, FULL_LINEAR and P12 have lower coordinate error than PRED_F; this larger-budget result is retained rather than replaced by a more favorable secondary cell. At 300 ps, errors approach the mean baseline, without establishing an irreducible predictability bound. TICA is evaluated here through the two NMSE targets, not an implied-timescale, Markov-state-model or equilibrium-kinetics criterion.

A21.4 Current-angle linear accessibility

Current-ϕ and current-ψ diagnostics use training-fitted affine readers of their sine/cosine pairs, evaluated on the same held-out trajectories and normalized by their respective training variances. They measure linear accessibility, not total information. The table gives lag 30 ps neural codes.

Method	d	Current ϕ NMSE	Current ψ NMSE
RECON	2	0.431083	0.113873
PRED_F	2	0.742717	0.128360
PRED_T	2	0.742100	0.127608
RECON	4	0.057790	0.021405
PRED_F	4	0.639320	0.116096
PRED_T	4	0.712298	0.113656

The worsening current-ϕ readout with better future prediction is an empirical trade-off in accessible features. It does not establish deletion of the angle information, minimal sufficiency, a discovered slow mode, or a molecular mechanism.

A21.5 Prespecified primary and conditional uncertainty

The paired differences are F loss of PRED_T minus PRED_F (Δ_F) and T loss of PRED_F minus PRED_T (Δ_T), after averaging the three seed losses at each test start. Each positive difference favors the objective trained for that target. A noncircular moving-block bootstrap samples contiguous paired-loss blocks with replacement, truncates the last sampled block to the original series length, and uses 10000 replicates. The main block is 5 ns; 1 and 20 ns are sensitivity lengths. Seeds are generated by NumPy SeedSequence from the fixed date seed 20261023, fold index and block length in ps. The intervals are percentile 95% intervals, conditional on these frozen fits and trajectories.

The prespecified robust cross-ranking requires both differences positive in every rotation, positive lower interval endpoints for both differences at every block length, and a training/validation dependence diagnostic. The latter requires the absolute autocorrelation of each torsion sine/cosine component at 5 and 10 ns to be at most 0.1. It passes in all rotations, but is a block-adequacy heuristic, not a proof of mixing or of interval coverage. The full cross-ranking criterion is not met. The saved intervals are given below; 5 ns remains the primary length regardless of sensitivity outcomes.

Table A9: Primary differences and saved moving-block intervals. Values are rounded for display; exact saved floating-point values are supplied in primary_comparison.json.

Fold	Block (ns)	Δ_F	F interval	Δ_T	T interval
0	1	0.000309	[0.000013, 0.000758]	0.000101	[-0.000167, 0.000370]
0	5	0.000309	[-0.000028, 0.000842]	0.000101	[-0.000158, 0.000380]
0	20	0.000309	[-0.000028, 0.000862]	0.000101	[-0.000157, 0.000419]
1	1	0.000390	[0.000026, 0.000895]	0.000257	[0.000013, 0.000572]
1	5	0.000390	[-0.000060, 0.001043]	0.000257	[0.000037, 0.000618]
1	20	0.000390	[-0.000062, 0.001141]	0.000257	[0.000030, 0.000644]
2	1	-0.000030	[-0.000081, 0.000025]	0.000214	[0.000020, 0.000558]
2	5	-0.000030	[-0.000089, 0.000026]	0.000214	[0.000019, 0.000578]
2	20	-0.000030	[-0.000090, 0.000010]	0.000214	[0.000021, 0.000594]

All main F intervals include zero; the third rotation has a negative F mean. Two main T intervals are positive and one includes zero. These outcomes do not confirm the prespecified primary, and do not establish equality, equivalence or target-independence. Rotation means and sensitivity cells do not replace that primary. Because trajectories exchange roles across folds, no independent-replication p-value is inferred by pooling folds. The intervals exclude uncertainty from drawing new training trajectories and are not population bounds or deterministic error certificates.

A21.6 Reproducibility coverage and boundaries

The small molecular supplement supplies source hashes, fixed configurations, analysis-source snapshots, verified numerical tables and a table-only arithmetic replay. The separately prepared `ALA2_primary_interval_replay_v1.zip` companion contains `paired_losses_fold0.npy` through `paired_losses_fold2.npy`, each a 249700-by-2 float64 array in test-start order. Columns are the three-seed mean differences (F loss of PRED_T minus PRED_F; T loss of PRED_F minus PRED_T); starts correspond to original frames 0 through 249699 at 1 ps spacing. It also contains the nine original arrays of bootstrap replicate means, primary_comparison.json, exact source hashes and `replay_primary.py`. The command python `replay_primary.py` –outdir REPLAY checks all input hashes and reproduces the existing means, all 10000 replicate means per fold/block length and the percentile intervals using the original seeds. This is arithmetic replay of already reported statistics; it performs no alignment, model inference, fitting or molecular dynamics. The companion remains separate from the small archive, which includes its member manifest and replay source. Neither it nor the generated model archive has yet been publicly deposited. The source trajectories are identified by upstream URL, hash and license. Full molecular training replay additionally requires

those trajectories and the separately retained model/generator materials; the primary-interval companion alone does not supply them.

These data add a fixed molecular benchmark within a known TAE setting. They do not test active sensing, kinetic rates, free energies, minimal molecular states or bit complexity. The field certificates in Supplementary Proofs SP0–SP4 do not apply here. Corrected E3 and every earlier sensing result retain their original status.

A22. Scientific descriptions and retained configuration labels

The main text uses scientific descriptions; the original machine-readable labels are retained without renaming the data or programs. Predictive versus present-state compression and the same-task mode-product diagnostic use D0; the critical-Brusselator horizon comparison uses D1; regime transfer uses D2. Query upgrades based on prediction disagreement use E0, and corrected Brusselator conditional ambiguity uses E3. Current-ambiguity spatial selection uses E4; future-score spatial selection uses E5; horizon transfer of selected sites uses E6. KS representation horizons use F0; its formal ambiguity allocation uses F1b; its formal future-score spatial selection uses F2. E1/E2 and F1a/F2a remain motivating or calibration/health analyses, respectively (A5), without separate policy confirmation. A19 and the configuration files preserve solver/data/model identities. This crosswalk changes names in the exposition, not experimental scope or evidential status.

A23. Frozen-code readout controls and external-cache validation

A23.1. Question, data roles and outcome history

The comparison tests what a richer reader can extract from an unchanged four-coordinate neural representation. It is not a new encoder or an information-sufficiency test. The original critical-regime cache contains 1,800 states in disjoint training/validation/test splits of 1,080/360/360. Only training states fit coefficients; validation chooses regularization. The original test was inspected before this diagnostic and remains exploratory. All controls predict the same 128-component raw future deviation field at horizon 10 and $B = 25/16$.

A separate 360-state cache was generated on 11 October 2026 under the same corrected 64-node, $\Delta t = 0.0025$ numerical pipeline and initial-state generator, with generation seed 202610110801. Its float32 inputs are unique and have no overlap with the original 1,800 inputs. Stored, unquantized float32 latents from the same three frozen PRED encoders are used directly. They are not reconstructed from the old 16-bit codewords. The stored float32 numerical futures are promoted to float64 for scoring. This is not an evaluation against a newly computed exact ODE/PDE future.

The external cache and old quantized PRED/RECON results were already known on 11 October; a true-ODE certificate for those coded pipelines followed. The physical-product diagnostic used the original cache on 20 October. The present nine readers were selected and frozen on 25 October before their external scoring. The external inputs and futures enter neither fitting nor validation. Nevertheless, full outcome-blind access/design history is not established. We describe a disjoint external-cache check, not new prospective confirmation. The prospective lattice outcomes in Part B have a separate, genuinely post-freeze generation history.

A23.2. Common readout objective and model selection

For raw features x_i, define $\tilde{x}_{ij} = (x_{ij} - m_j)/s_j$ using training-only m_j and $s_j = \max(\mathrm{std}_{\mathrm{train}}(x_j), 10^{-12})$. Expand the raw code before this column scaling. The PRED quadratic dictionary is $(z_1, \dots, z_4, (z_p z_q)_{p \le q})$, four coordinates and ten products. The physical dictionary is the six amplitudes from A20 with their 21 products. The full-current reference uses 128 raw current-field entries.

For training future mean m_Y, use a single global centered RMS s_Y. With $\widetilde{Y}_i = (Y_i - m_Y)/s_Y$, the coefficients solve

$$\min_{W,c} \sum_{i \in \mathrm{train}} \|\widetilde{Y}_i - c - \tilde{x}_i W\|_2^2 + \alpha \|W\|_F^2, \qquad \alpha \in \{0, 10^{-6}, 10^{-4}, 10^{-2}, 1\}.$$

The intercept c is unpenalized. The zero-ridge candidate uses minimum-norm SVD least squares; positive ridge uses an augmented SVD least-squares formulation. Raw-field predictions are $m_Y + s_Y(c + \tilde{x}W)$. Validation minimizes unrounded mean per-state relative L2, with exact ties resolved toward larger α. No validation/test target enters scaling. All nine selected coefficient sets and scalers were sealed before new test evaluation; no post-score choice, recalibration, saturation filtering or target-norm filtering is made.

Table A10: Selected reader regularization and coefficient budgets.

Reader	Columns	Selected α	Reader coefficients
PRED affine / seed 04	4	0	640
PRED quadratic / seed 04	14	0.01	1920
PRED affine / seed 05	4	0.01	640
PRED quadratic / seed 05	14	1	1920

Reader	Columns	Selected α	Reader coefficients
PRED affine / seed 06	4	0.01	640
PRED quadratic / seed 06	14	0.01	1920
Physical affine	6	1	896
Physical quadratic	27	0	3584
Full-current affine	128	1	16512

The 25,028-parameter neural encoder remains fixed and is additional to the reader. Physical maps instead use known species/spatial directions. Base coordinate counts are 4 versus 6; expanded columns are 14 versus 27. Centered affine-output rank upper bounds are the column counts, not measured ranks. Common target/column scaling, ridge grid and selection criteria remove those reader-fitting differences, but do not equalize prior, representation training, feature count or output rank. Historical A20 reader coefficients and table entries remain separate: the new affine PRED mean 0.038800 is not a correction of its original 0.038779.

A23.3. Main and secondary metrics on both banks

PRED metrics average the separately scored three seeds. Relative/absolute quantiles are computed after averaging the three per-state losses; no averaged prediction field is constructed. Means of squared losses are averages of seed-specific squared losses, not squares of seed-averaged losses.

Table A11: Forecast metrics on the original test and previously examined external cache.

Reader / bank	Rel. L2	Abs. L2	Rel. median	Critical R^2	Uniform R^2
Full-current affine / original	0.151973	0.01396506	0.106114	0.977387968	0.988383897
Physical affine / original	0.150564	0.01392634	0.104222	0.977351986	0.988691608
PRED affine / original	0.038800	0.00350691	0.029120	0.999672877	0.996634985
PRED quadratic / original	0.026861	0.00223297	0.020685	0.999685707	0.998978008
Physical quadratic / original	0.022884	0.00196525	0.014037	0.999546854	0.999806017
Full-current affine / external	0.172576	0.01481301	0.113002	0.972741761	0.990631244
Physical affine / external	0.168710	0.01476141	0.118657	0.972617282	0.990823343
PRED affine / external	0.041683	0.00347176	0.029254	0.999586469	0.996380512
PRED quadratic / external	0.029014	0.00229040	0.021038	0.999585011	0.998943833
Physical quadratic / external	0.024310	0.00200735	0.014811	0.999444281	0.999797814

Table A12: Relative-error quartiles and mean squared relative losses on both banks.

Reader / bank	Rel. Q1	Rel. Q3	Mean squared rel. L2
Full-current affine / original	0.056078	0.213251	0.042280855
Physical affine / original	0.056045	0.205183	0.042148766
PRED affine / original	0.020182	0.044856	0.002522187
PRED quadratic / original	0.013273	0.033157	0.001259272
Physical quadratic / original	0.007743	0.027450	0.001288753
Full-current affine / external	0.071808	0.218021	0.061907455
Physical affine / external	0.073770	0.221263	0.057719684
PRED affine / external	0.020299	0.047789	0.003844585
PRED quadratic / external	0.014696	0.034911	0.001914781

Reader / bank	Rel. Q1	Rel. Q3	Mean squared rel. L2
Physical quadratic / external	0.008643	0.029233	0.001405475

On the original test, quadratic PRED has smaller mean squared relative error than physical quadratic readout, although larger mean relative L2. On the external cache, both means favor the physical reader. The squared-loss ranking is therefore not stable across these banks. Absolute-L2 means retain the mean-relative-L2 ordering. These differences preclude a universal winner across losses.

A23.4. Paired descriptive uncertainty and per-seed results

The three contrasts below, 10,000 paired initial-state bootstrap resamples, seed 202610250902 and linear percentile interpolation were fixed before external scoring. The same resampling indices are used across contrasts. Seeds are averaged within an initial state before resampling; 1,080 model–state pairs are not treated as independent observations. A positive difference is reference loss minus candidate loss.

Table A13: Prelisted external-cache contrasts and descriptive individual intervals.

Candidate versus reference	Mean gain	Better/worse	Individual 95% interval
PRED quadratic vs affine	0.012669	274/86	[0.010602, 0.014919]
Physical quadratic vs affine	0.144400	353/7	[0.129052, 0.161440]
Physical vs PRED quadratic	0.004704	233/127	[0.002482, 0.006910]

Nominal three-contrast Bonferroni endpoints use percentiles $0.05/(2 \times 3)$ and its complement, corresponding to nominal 98.33% intervals per contrast:

Table A14: Nominal multiplicity-adjusted descriptive endpoints for the same three contrasts.

Contrast	Nominal adjusted endpoints
PRED quadratic vs affine	[0.010197, 0.015422]
Physical quadratic vs affine	[0.125347, 0.166380]
Physical vs PRED quadratic	[0.001999, 0.007400]

All three point gains and both sets of endpoints are positive, but these are approximate descriptive percentile intervals, conditional on the frozen pipelines. Nominal endpoint adjustment does not supply finite-sample simultaneous-coverage guarantees. They do not account for training, hyperparameter selection, earlier outcome access or solver error, and introduce no confirmatory primary or new p value. No equal-effect or per-state dominance claim is made; all reverse-direction states are retained, with no ties in these three contrasts.

Table A15: Individual fixed-seed errors on both banks.

Seed	Original affine	Original quadratic	External affine	External quadratic
20261004	0.040392921	0.028146597	0.041840648	0.028346196
20261005	0.036533581	0.025912887	0.039887028	0.028914731
20261006	0.039472839	0.026524441	0.043321714	0.029782498

Each seed's mean improves with the quadratic PRED reader on both banks. Physical quadratic readout is lower in mean relative L2 than each PRED quadratic seed. Critical-amplitude R^2 need not improve with the richer neural reader: on the external cache it decreases for two seeds and increases for one. Field-error improvement does not imply monotone component improvement.

A23.5. Complete nonorthogonal component accounting

Using A20's critical, fast, uniform and remaining-field maps, the error field is decomposed before squaring. The tables report the mean of each squared component divided by $(\|Y_i\| + 10^{-10})^2$ and all six doubled cross terms with that same denominator. Selected diagonal terms alone are not additive error fractions; this identity decomposes mean squared relative loss, not mean relative L2.

Original test.

Table A16: Complete nonorthogonal squared relative-error decomposition on the original test.

Term	PRED affine	PRED quadratic	Physical quadratic
critical	0.000697989926	0.00069531177	0.00121030019
fast	2.1017857e-06	3.56713897e-08	1.6329376e-07
uniform	0.000991404278	0.000425267868	5.4623649e-05
remaining	0.000797479351	0.000133247446	2.32922194e-05
critical / fast	1.68195816e-06	1.71877815e-07	6.9473114e-08
critical / uniform	8.24158771e-21	2.71670837e-21	-3.43462937e-21
critical / remaining	1.53532611e-08	1.17750553e-08	5.11740358e-08
fast / uniform	-8.14809395e-23	-1.20205807e-23	-2.43806984e-23
fast / remaining	-7.92683122e-10	8.89305545e-10	-7.09317503e-11
uniform / remaining	3.15146909e-05	5.22491292e-06	2.53010897e-07
Total	0.00252218655	0.00125927221	0.00128875294

External cache.

Table A17: Complete nonorthogonal squared relative-error decomposition on the external cache.

Term	PRED affine	PRED quadratic	Physical quadratic
critical	0.00113819553	0.0011131581	0.00132380096
fast	2.03535313e-06	4.42451832e-08	2.11945623e-07
uniform	0.0013340834	0.000599006421	5.59239131e-05
remaining	0.0013186727	0.00019358704	3.00026175e-05
critical / fast	2.0505769e-06	-5.03081018e-07	-4.55385287e-06
critical / uniform	7.89383125e-21	-1.10275858e-21	-7.1080415e-22
critical / remaining	1.53279783e-08	4.05678908e-10	-4.61835008e-09
fast / uniform	-1.41969502e-24	-1.26003067e-23	5.61182817e-24
fast / remaining	-4.23231877e-10	1.09034467e-09	-1.58882216e-10
uniform / remaining	4.95321927e-05	9.48658748e-06	9.44696209e-08
Total	0.00384458466	0.00191478081	0.00140547528

The physical quadratic model has smaller uniform and remaining-field squared errors on both banks, while PRED retains more accurate critical amplitudes. This is a localization of saved forecast

errors, not a unique causal attribution to a quadratic physical mechanism. The richer PRED reader reduces part of the gap without new observations; it does not prove that no task-relevant information was lost by encoding.

A23.6. Readout provenance, technical retry and access

Nine reader instances and five ridge values define 45 unique configurations. The initial numerical execution stopped at the 41st solve, the full-current zero-ridge candidate, before new test scoring. A Gram-form stationarity check was unreliable for near-collinear float32-derived features with large minimum-norm coefficients. The technical retry used augmented SVD for positive ridge and a direct residual backward-error diagnostic; it preserved the training objective, data, feature maps, ridge grid and validation rule. The unregularized full-current candidate was not selected. Unsaved intermediate coefficients required rerunning the same grid: 86 actual solve calls (41 plus 45), not 45 total machine calls. Both failed and completed run records are retained. No test-dependent scientific protocol change was made.

The actual raw readout array SHA256 is `b41e175c25a39242fe3b6fddcbe36b1ea3ec66844c5f4acabd1244decde03c5d`. Small result/configuration manifests identify the original cache, frozen neural artifacts, selected coefficient file and nine-reader external outputs. Reader selection was sealed at 00:46:38 +08:00 on 25 October 2026; the external scoring plan was sealed at 01:04:07 +08:00. No target generation, neural execution or fitting was performed during external evaluation or manuscript integration.

The two numerical batches retain their original resource receipts: the completed grid retry took 1.54 s and peaked at 73.72 MiB; the first-start-to-retry-completion interval, including technical handling, was 148.24 s. The external evaluation kernel took 0.554 s and peaked at 48.61 MiB, excluding preparation, reporting and arithmetic replay. These timings are execution provenance, not fair speed comparisons between algorithms. Both used one CPU affinity and one BLAS thread. Prior checks reconstructed predictions from raw-unit coefficients and DCT-I projections and replayed losses/interval arithmetic without fitting; their receipts are retained and do not constitute a new physical replication.

The evaluation-level field arrays, selected readers and no-fit numerical replay are supplied in Online Resource 3 (`Evaluation_Replay.zip`, field directory). The companion contains the two evaluated sets, not the complete training cache, and has not been added to the fixed public Zenodo version cited in this manuscript. The earlier finite-task proofs and true-ODE certificates concern the old frozen coded models; no strict true-ODE risk interval was calculated for these nine new unquantized readers. Their target enclosures are not treated as a new ranking certificate. E3 remains solver-corrected, primary not confirmed, $p = 0.084983$.

A24. Historical-input dependence in recorded electric-drive signals

A24.1 Task, data identity and fixed comparison

We use Coupled Electric Drives (CED), an established measured-signal benchmark whose report describes two motors, a flexible belt and a spring-supported pulley [38]. Its output sensor is insensitive to velocity sign. Here $u[k]$ denotes the released control input, and $z[k]$ the released measured output, rather than the unobserved signed velocity. Results use the released output scale. Five existing 500-sample records came from one pinned third-party mirror; byte identity with the original author archive remains unresolved. The report's nominal clock does not independently verify timing or calibration in this mirror.

Development uses `DATAPRBS u1/z1` and `u2/z2`, and `DATAUNIF u11/z11`; validation uses `DATAPRBS u3/z3`, and testing `DATAUNIF u12/z12`. No record is split between groups. With zero-based indexing, origins are $k = 5 + 11j$, $j = 0, \dots, 44$, giving 135 development and 90 evaluation origins. Full supports $[k-5, k+5]$ do not overlap within a record. This prevents direct reuse of another origin's target as an output feature; it does not establish independent errors or independently generated waveforms. Existing checks find scaled/shifted correspondence between input records on their analyzed support.

The target is $z[k+5]$. Both learned predictors receive current output $z[k]$ and ten controls $u[k-5], \dots, u[k+4]$. These include recorded future controls: the task is conditional replay, not unknown-forcing free-run prediction. Predictor S uses these shared inputs; H additionally uses the history difference $d[k] = z[k] - z[k-5]$. Each predicts the increment from the current output, using a separate development-only ridge fit with population-standard-deviation normalization, residual-sum-of-squares objective, fixed $\lambda = 1$ and an unpenalized intercept. Validation did not tune or refit either model. Since shared coefficients are refitted, H versus S does not isolate the varying history term.

The third arm H0 keeps H's coefficients, intercept, normalization, current output and controls unchanged. It substitutes the original development mean $\mu_d = 0.15494957717179916$ for $d[k]$, making H's standardized history coordinate $h = (d[k] - \mu_d)/s_d$ zero; s_d is the development standard deviation. Writing H's increment prediction as $b_H + w_H^T x_{\text{std}} + v_H h$, where x_{std} contains the shared standardized inputs, the replacement changes the prediction by $\delta = -v_H h$. It inherits H's history-dependent training and is not another current-only fit. This distinction concerns a fixed model's input dependence, as opposed to importance across fitted models [39]. Our deterministic development-mean replacement is not the permutation-based model-reliance estimator in Ref. [39].

A24.2 Complete recorded outcomes

Table A18 retains both records, four error metrics, persistence P0 (predict $z[k]$) and the zero-fit trend P1 (predict $z[k] + d[k]$). Error is prediction minus target. The original S/H comparison was followed by the H0 diagnostic after its outcomes were read; H0 is post-hoc, not a new blinded confirmation. All entries below are recomputed only from saved predictions. No new fit or forecast was produced for this integration.

H has lower RMSE than S on both records, while persistence already has lower test RMSE than H. For the same H, mean replacement increases validation RMSE but decreases test RMSE. The corresponding H0-minus-H mean-squared-error changes are $+0.012435090$ and -0.064930474. Lower/higher squared-error counts are 19/26 on validation and 29/16 on testing; all origins are retained. With original H error e, the standard expansion is

$$\overline{(e+\delta)^2 - e^2} = 2\overline{e\delta} + \overline{\delta^2}. \tag{11}$$

This accounts for the existing errors; it is not a new theorem or target-based bias correction. The stored prediction changes agree with removal of the development-centered history term to maximum residual below 6×10^{-16}. A fixed model's error decreasing under information removal is an

Record	Predictor	RMSE	MAE	Mean error	Max. absolute error
Validation	H	0.297689	0.241137	0.185493	0.634636
	H0	0.317889	0.263366	0.149442	0.767620
	S	0.380470	0.328053	0.287589	0.828373
	P0	0.402563	0.322019	0.031366	0.967033
	P1	0.731715	0.565025	0.057523	1.816850
Test	H	0.666120	0.513162	0.390970	1.784248
	H0	0.615456	0.468293	0.328337	1.528974
	S	0.707656	0.569937	0.502963	1.575840
	P0	0.643946	0.502618	−0.069570	1.811966
	P1	0.838660	0.662163	−0.138380	2.358974

Table A18: CED saved-prediction audit, 45 origins per evaluation record, in released output units. H and S are separately fitted; H0 changes only the varying history input in the same H. P0 and P1 use different input/estimation budgets. No independent-origin confidence intervals are attached.

established possibility [39]; this case gives the two record-specific directions for the declared predictor.

A24.3 Observation budget and interpretation

H requires two output readings per origin and H0 one: 180 versus 90 logical readings across the evaluated origins. Both use the same 900 control values and retain H's development history cost. These are observation dependencies, not measured hardware energy or acquisition savings. The input replacement need not lie on the naturally occurring joint input support. It does not estimate optimal current-information risk, intrinsic historical information, or an intervention on the apparatus. The sensor, noise, correlated controls and limited regression family are not separately identified causes of the loss change.

The companion `CED_Evaluation_Replay.zip` contains the saved per-origin predictions/targets, frozen parameters, complete metrics, costs, a manifest and a standard-library arithmetic replay. It verifies the original H scores before computing the contrast and checks all 90 historical-term identities. It does not execute H, refit models, reconstruct raw acquisition or supply a new public accession. Original training and replacement scripts, execution freezes and seal records remain in the source audit; application-level access separation is not an operating-system security proof. These previously examined records are a supporting fixed-predictor diagnostic, not another replication or information certificate for the two main systems.

A25. Prospective field validation and common-output-space controls

A25.1. Frozen scope and data identity

This experiment tests the selected readers on new initial states after all choices were sealed. It is separate from the reused original test and the previously examined external cache in A23. The corrected Brusselator source, $B = 25/16$, 64 nodes, $L = 2\pi$, $\tau = 10$, $\Delta t = 0.0025$, initial-mode distributions and positivity shrinkage are unchanged. Exactly 720 independent initial states are generated with fixed seed 202610290801; no state is replaced according to prediction error. Their Float32 current-field rows are unique and have no byte overlap with the original 1,800 states or historical 360 external states. Raw Float64 physical initial conditions and endpoints are retained. The primary target follows the original Float32 deviation-field convention, promoted to Float64 for loss evaluation, with norm denominator offset 10^{-10}.

The old nine reader instances retain their original coefficient bytes. Three saved neural encoders perform CPU forward passes only. On the old training and validation inputs, the largest raw-latent difference from saved artifacts is $2.384185791015625 \times 10^{-6}$, below the prespecified 10^{-4} tolerance; all six seed–split checks pass but are not bitwise identical. No neural training, new architecture, parameter search on new outcomes or repeat generation is performed.

A25.2. Common output spaces and selection

One SVD of the original 1,080 training future fields, centered by their training mean and divided by one global centered RMS, provides nested POD bases with 4 and 14 columns. Each singular-vector sign is fixed by making its largest-absolute entry positive. No validation or test target selects the basis. Each existing reader family predicts coefficients in a shared basis and reconstructs the 128-component field with the same training mean and scale. Feature columns, products, train-column standardization floor 10^{-12}, unpenalized intercept, ridge grid and validation selection follow A23. There are $2 \times 9 \times 5 = 90$ additional fits and 18 selections, in addition to the unchanged old nine readers; all 27 pipelines are evaluated on the same 720 states.

The only selection metric is the original 360 validation states' unrounded mean relative L2 of the reconstructed full field; exact ties choose larger ridge. All 90 fitted coefficient arrays and all validation values are retained. The 18 selections were sealed before initial-state generation. The common-output restriction does not equalize feature columns, parameters, known physical modes or encoder costs. No standard reduced-rank-regression optimum is claimed.

A25.3. Primary and complete pipeline summaries

The sole prespecified primary is the unrestricted PRED affine–quadratic comparison. Losses are first averaged over the same three fixed model seeds within each initial state, then averaged across 720 initial states. The 2,160 model–state entries are not independent physical samples, and model predictions are not averaged before loss evaluation. With bootstrap seed 202610290802, 10,000 paired initial-state resamples define a percentile interval with linear quantile interpolation. The prespecified support criterion is a strictly positive lower endpoint. This is an approximate conditional interval, not finite-sample guaranteed coverage, a true-ODE certificate or a new p-value. It excludes neural-training, reader-selection and numerical-solver uncertainty. Other reader, rank, physical-component and loss comparisons remain preplanned descriptive diagnostics.

The mean affine-minus-quadratic loss difference is 0.009919621, interval $[0.008859427, 0.011014209]$. The lower endpoint is positive. Relative mean error decreases 26.83%; the quadratic reader improves 544 states and worsens 176, with no ties.

Table A19: All saved reader metrics on the 720 prospective initial states.

Output	Reader	Mean rel. L2	Mean abs. L2	Critical R^2	Uniform R^2
Free	PRED affine	0.036969580	0.003227854	0.999550783	0.997620850
Free	PRED quadratic	0.027049959	0.002270451	0.999560417	0.999041013
Free	Physical affine	0.165096608	0.015252022	0.966537888	0.991283770
Free	Physical quadratic	0.023878336	0.002083126	0.999399129	0.999791831
Free	Full-current affine	0.168151147	0.015383750	0.966322806	0.991164733
POD4	PRED affine	0.036969658	0.003227873	0.999550798	0.997620851
POD4	PRED quadratic	0.027101026	0.002278092	0.999562492	0.999041014
POD4	Physical affine	0.165096498	0.015252035	0.966540654	0.991283783
POD4	Physical quadratic	0.023939583	0.002091625	0.999400547	0.999791835
POD4	Full-current affine	0.168150468	0.015383718	0.966325437	0.991164740
POD14	PRED affine	0.036969580	0.003227854	0.999550783	0.997620850
POD14	PRED quadratic	0.027049959	0.002270451	0.999560417	0.999041013
POD14	Physical affine	0.165096608	0.015252022	0.966537888	0.991283770
POD14	Physical quadratic	0.023878336	0.002083126	0.999399129	0.999791831
POD14	Full-current affine	0.168151147	0.015383750	0.966322806	0.991164733

Neural rows average individual-model metrics. They are not ensemble-field forecasts. All relative-error norms refer to the same full field.

Table A20: Individual fixed-seed errors under the three output restrictions.

Seed	Free affine	Free quadratic	POD4 affine	POD4 quadratic	POD14 quadratic
20261004	0.038648966	0.028093625	0.038649576	0.028142017	0.028093625
20261005	0.034978167	0.025896197	0.034977968	0.025948113	0.025896197
20261006	0.037281606	0.027160054	0.037281428	0.027212948	0.027160054

Each fixed seed improves in mean in each output restriction. Under POD4, the seed-loss-averaged relative reduction is 26.69%, with 543 improved states and 177 worsened. The unrestricted physical quadratic reader has lower mean than every neural quadratic seed, with 11.73% lower mean relative L2 than the neural quadratic aggregate; 453 states favor the physical reader and 267 the neural reader. No separate confirmatory test or interval is assigned to these secondary comparisons.

A25.4. Error distributions and physical components

Table A21: Per-state relative-error distribution summaries on the prospective bank.

Output	Reader	Q25	Median	Q75	Q90	Maximum
Free	PRED affine	0.018717449	0.027576656	0.045285558	0.071649884	0.171836675
Free	PRED quadratic	0.013766712	0.020511698	0.033335227	0.054574930	0.126245123
Free	Physical quadratic	0.008305226	0.015633156	0.030023469	0.053004578	0.248750627
POD4	PRED affine	0.018715581	0.027576687	0.045286315	0.071649310	0.171836437
POD4	PRED quadratic	0.013802110	0.020576607	0.033386142	0.054576435	0.126246056
POD4	Physical quadratic	0.008377721	0.015674791	0.030029987	0.053005106	0.248750736
POD14	PRED affine	0.018717449	0.027576656	0.045285558	0.071649884	0.171836675
POD14	PRED quadratic	0.013766712	0.020511698	0.033335227	0.054574930	0.126245123

Output	Reader	Q25	Median	Q75	Q90	Maximum
POD14	Physical quadratic	0.008305226	0.015633156	0.030023469	0.053004578	0.248750627

The following unrestricted error decomposition uses A20/A23's critical, fast, uniform and remaining-field maps. Each diagonal or doubled cross term is averaged after division by the squared per-state target norm plus offset. This identity decomposes mean squared relative loss; it is not a percentage attribution of mean relative L2. Neural component terms average the individual-model squared-loss terms.

Table A22: Complete nonorthogonal squared relative-error decomposition on the prospective bank.

Term	PRED affine	PRED quadratic	Physical quadratic
critical	0.0006953105051	0.0006908924595	0.001148589429
fast	1.9683726e-06	3.391688672e-08	2.002289073e-07
uniform	0.000869406934	0.0004187805103	5.360957293e-05
remaining	0.0006376387756	0.0001176552168	2.464327542e-05
critical / fast	1.679537062e-07	3.618970176e-08	-1.905163705e-06
critical / uniform	2.968588139e-21	1.486002253e-21	-8.902715995e-22
critical / remaining	2.927410445e-09	-1.448898311e-09	6.798019654e-09
fast / uniform	-3.844605283e-23	-5.417266918e-24	7.670483938e-24
fast / remaining	-2.018963805e-10	8.00748579e-10	-3.539529904e-10
uniform / remaining	2.617783159e-05	5.268689241e-06	1.050473854e-07
Total	0.002230673098	0.001232666334	0.001225248834

The physical quadratic reader has smaller uniform and remaining-field errors while the neural quadratic reader has more accurate critical amplitudes. These are retained component diagnostics, not a unique causal decomposition of the physical or neural computation. All 27 individual pipeline metrics, all four components and six cross terms, absolute-error quantiles and complete per-state losses are present in the retained result arrays/JSON, including reverse directions.

A25.5. Numerical ranks, coefficient budgets and projection reference

Numerical rank counts coefficient-map singular values above 10^{-10} times the largest singular value, with the threshold applied in the declared standardized-feature convention. All singular values and absolute thresholds are retained. The table reports actual computed ranks, not only algebraic allowances. Coefficient counts include the intercept but exclude the common basis and target mean/scale.

Table A23: Numerical ranks and reader coefficient budgets for the shared output spaces.

Reader	Columns	Rank4	Rank14	Coeff4	Coeff14	Ridge4 / Ridge14
Neural A-4	4	4	4	20	70	0 / 0
Neural Q-4	14	4	13	60	210	0.01 / 0.01
Neural A-5	4	4	4	20	70	0.01 / 0.01
Neural Q-5	14	4	13	60	210	1 / 1
Neural A-6	4	4	4	20	70	0.01 / 0.01
Neural Q-6	14	4	14	60	210	0.01 / 0.01
Physical affine	6	4	6	28	98	1 / 1

Reader	Columns	Rank4	Rank14	Coeff4	Coeff14	Ridge4 / Ridge14
Physical quadratic	27	4	14	112	392	0 / 0
Full-current affine	128	4	14	516	1806	1 / 1

The common bases carry 128×4 or 128×14 entries, shared by all readers in their group, plus the shared target normalization. Each neural encoder retains its separate 25,028 trained parameters. Four neural coordinates expand to 14 columns; six known physical coordinates expand to 27. At POD4, all nine maps have numerical rank four, yet neural affine and quadratic readers still have 20 versus 60 coefficients, and the physical quadratic reader has 112. At POD14, the neural affine ranks remain four, while neural quadratic ranks are 13, 13 and 14. Shared output allowance is not matched parameter count, physical prior, input information or computational cost.

Projecting the new true future onto the training POD4 affine space gives mean relative L2 0.001086462; POD14 gives $2.32996491 \times 10^{-8}$. This uses future truth and is an output-truncation reference, not an implementable predictor. POD14 retains essentially the numerical future field, so its near-agreement with unrestricted results is not another independent replication.

A25.6. Fixed numerical sensitivity

The fixed first 36 initial states receive the prespecified half step $\Delta t = 0.00125$. Float64 future-deviation endpoint relative changes have mean $2.43815042 \times 10^{-5}$ and maximum $6.03478405 \times 10^{-5}$. Their primary mean difference changes from 0.010948234 to 0.010947121, with the same positive direction. Using Float64 rather than the primary Float32 deviation targets on all 720 changes the primary mean by $8.01961362 \times 10^{-11}$.

These diagnostics are restricted to the selected step-size comparison and target precision. They are neither certified error bounds for all 720 states nor a continuous-PDE assertion. The old finite-task/true-ODE proofs in the separate attachment retain their original quantized models and sample sets and do not certify these new unquantized readers. E3 remains solver-corrected, primary not confirmed, $p = 0.084983$; the new representation experiment does not alter that sensing result.

A25.7. Ordering, archive and evaluation replay

The local execution protocol, sole primary, random seeds and budgets were frozen first. The 90-fit/18-choice seal precedes the initial-state record, which precedes future completion; every frozen reader and generated-array hash was verified before scoring. No method-level error on the new bank was examined before all 720 main futures and fixed 36 half-step futures were saved. This is a prospective local freeze, not a public preregistration. The original result files and versioned manuscript remain unchanged.

The accompanying anonymized `Prospective720_Evaluation_Replay.zip` (Online Resource 5) contains the original fitting cache and saved latents/scalers, three frozen neural checkpoints, original nine and new 18 reader parameters, all 90 candidate coefficients and validation values, training POD, physical initial conditions and Float64 endpoints, Float32 targets, new raw latents, all 27 forecasts/losses, primary bootstrap draws, evaluation code and manifests. It is an evaluation copy of the retained result archive: scientific arrays and frozen parameters retain their original bytes, while identifying metadata and local paths are sanitized. Historical provenance records retain the original freezing and generation order; their historical source hashes identify the original source bytes, whereas the accompanying manifest identifies the anonymized copies. Provenance-only execution sources are marked separately from the runnable evaluation replay. The companion is provided with the submission and is separate from the fixed public Zenodo record; no new public accession is claimed.

After extraction, `python -X utf8 replay_bundle.py` verifies all member hashes, the 27 forecasts and per-state losses, input nonoverlap, and the saved primary bootstrap using independent

arithmetic expressions; it performs no training, fitting, encoder execution or integration. A separate retained local arithmetic check also verifies all 90 validation residuals/losses, all 18 choices and the training-POD eigen-equations. Prediction reconstruction uses raw-unit coefficients and DCT-I physical amplitudes; norms use compensated summation and bootstrap replay uses count vectors. These checks are evaluation-level arithmetic agreement, not independent training or independent-team replication.

The 90-fit, generation, evaluation and arithmetic-replay stages took 0.196, 7.497, 1.326 and 0.873 s, respectively, in the retained one-core/one-BLAS-thread execution. The largest stage peak working set was 227.38 MiB. These are local execution receipts, excluding code preparation and review, not fair algorithmic speed benchmarks. No additional experiment is run during manuscript integration.

Part B. Frozen prospective observational refinement

B1. Lattice energy, static calibration and driven protocol

B1.1 Static model and interpretation of a_c

The full square-lattice energy used for the present finite-size experiment is

$$F[q, y; a] = \sum_i \left(\tfrac{1}{2}aq_i^2 + \tfrac{1}{4}uq_i^4 + \tfrac{1}{2}\kappa y_i^2 - gq_i y_i\right) + \tfrac{1}{2}K\sum_{\langle i,j\rangle}(q_i - q_j)^2.$$

The present configuration has u = K = κ = g = T = Δx = Γ = 1, with periodic boundary conditions. Completing the square gives

$$\tfrac{1}{2}\kappa y_i^2 - gq_i y_i = \tfrac{1}{2}\kappa(y_i - gq_i/\kappa)^2 - \tfrac{1}{2}(g^2/\kappa)q_i^2.$$

Thus the equilibrium q marginal has quadratic coefficient r = a - g^2/κ = a - 1; y|q is Gaussian with mean gq/κ and variance T/(κΔx^2). The finite-temperature critical calibration was performed in r coordinates before converting back to a; r = 0, or a = 1, is a *mean-field quadratic reference*, not the finite-temperature, finite-size transition estimator.

The archived equilibrium calibration sampled 49 parameter/size settings with four independent chains per setting (196 chains), using L = 32, 64 and 128. The L32 coarse scan preceded refined and common-bracket runs; the refined/zoom chains used 2,000 burn-in sweeps, 8,000 production sweeps and thinning by 5. Heat and cold starts alternated. Binder's U4 = 1 - $\langle m^4\rangle/(3\langle m^2\rangle^2)$ and other observables were saved. Two finite-size crossing estimates were -0.2685684 for L32/L64 and -0.2732469 for L64/L128. Assuming a *fixed* crossing shift L^(-p), illustrative choices p = 2, 3, 4 gave extrapolants -0.2748065, -0.2739153 and -0.2735588, respectively. For p = 3 the archived bracket/bootstrap sample reported [-0.2774175, -0.2693472]. These are conditional numerical statements and do not account for unknown finite-size corrections or all discretization uncertainties. The physical evaluation used the frozen protocol reference a_c = -0.274.

Source: `equilibrium_critical_calibration.md` (archived static calibration). The archived calibration is a completed control, not a new experiment in this revision.

B1.2 Protocol and forcing intervals

The actual driven control is

$$a(t) = a_c + \max[-1, \min(1, -t/\tau_Q)], \qquad \tau_Q = 64,$$

with nominal crossing reference t = 0. The physical future begins at the saved complete state of age t and ends 16 model-time units later. Values at the nominal endpoints (the actual integrator samples intermediate midpoints) are:

Table B1. Driving windows and control-parameter endpoints in dimensionless time.

Starting age t	End age t+16	a(t)	a(t+16)	Window
-8	8	-0.149000	-0.399000	Crosses nominal reference
0	16	-0.274000	-0.524000	Begins at reference
16	32	-0.524000	-0.774000	Driven later window
20	36	-0.586500	-0.836500	Driven later window

All windows remain on the ramp. They are **not** common-future-law comparisons at a fixed value of a; starting age also changes the terminal class distribution and the fixed age-conditioned prediction lookup. The recorded times are dimensionless model units. The finite-size calibrated zero is not a separately certified sharp thermodynamic critical crossing.

B2. Executed stochastic update and available numerical controls

B2.1 Formal continuous target and executed discrete kernel

The intended overdamped order-field and auxiliary relaxational dynamics are consistent with the force from F, with Γ = 1 and ν = 1. For the present reproducibility claim, however, the **executed** discrete Markov kernel is the scientific object. The original immutable physical source is `source/quench_pilot_101.py`, SHA256 `043e7d715b2375fec918d3ea8d4dceaa106fa3afc91b6fe0c3e5ced4fef1b912`, invoked by `101_stage9_O16_gated_future_runner.py`, SHA256 `8ee983d66e80e67564ea6647a4107eced67d7c8b1e2cce21b5ce8fd7a44d52ac`. Parameters: L = 64, ν = 1, τ_Q = 64, Δt = 0.02, four 200-step segments and periodic nearest neighbors. The original `advance()` routine evaluates the control at t_mid = start + (step + 0.5)Δt, then executes:

1. A conditional OU half-step for every y site, holding q fixed. Let d = exp(-νΔt/2), s = sqrt(1 - d^2). With the present unit T and κ, this is y ← d y + (1 - d)q + sξ_y, with independent standard-Gaussian ξ_y.
2. A checkerboard parity-0 then parity-1 local Metropolis-adjusted q proposal at each q site. With nearest-neighbor sum N_i, h_i = N_i + y_i, b = a(t_mid) + 4 and f(x) = h_i - bx - x^3, propose x′ = x + Δt f(x) + sqrt(2Δt)ξ_q. For site-local potential difference ΔU = b(x′^{2}-x^2)/2 + (x′^{4}-x^4)/4 - h_i(x′-x), the log acceptance ratio is

$$\log A = -\Delta U + \frac{(x' - x - \Delta t f(x))^2 - (x - x' - \Delta t f(x'))^2}{4\Delta t}.$$

3. Another conditional OU half-step with the *updated* q field. If a q proposal is rejected, that site retains its old value but the model-time step still elapses. The four segment boundaries do not reset the control time or the initial state.

The executed q update is a Metropolized discretization. Literature on other Metropolized SDE schemes [33] cannot be used as a free-standing theorem that this time-dependent split scheme converges pathwise to an unspecified continuum law. The primary numerical findings are defined for the archived discrete kernel and declared Δt; they do not depend on asserting continuum-time exactness.

B2.2 Archived controls, separately from the present 32,768 futures

The archived `numerical_controls.md` records, among others: Gaussian Fourier one-step error 6.4×10^{-20} and lattice-noise variance 0.1612667 versus target 0.16 in an earlier control; the replaced older Euler/implicit-quartic proposal showed q^2 biases about 4% at Δt = 0.02 and 2% at Δt = 0.01 and is **not the future-generation entry point here**. The corrected local-MALA/OU kernel was checked against fixed-a Gaussian moments at ν ∈ {0.03, 1, 10} and Δt = 0.02/0.01; a Gaussian q^2 reference 0.2540498788 was compared with selected observed means around 0.2530–0.2557. Nonlinear fixed-a runs at a = 0.726 and -0.274, L16, several ν and two Δt values, had archived acceptance fractions 0.99041–0.99706 and sampled moment checks compatible with selected static references.

Other archived checks include effective small-L physical relaxation slopes, finite-temperature coarsening checks and an isolated curvature control. These are **limited numerical consistency**

tests; they are not proof of a thermodynamic critical dynamic exponent, full finite-Δt error convergence for the present nonstationary 64×64 protocol, or full equilibrium-parent sampling. Those historical acceptance fractions are **not** the acceptance distribution of all 32,768 new future paths. The raw future NPZ contains its own four-segment acceptance array, and any summary of it would have to be separately computed and labelled. No such new estimate was introduced here.

B3. Four observed sensors, S14 metadata, quantization and tail rules

B3.1 Exact original selection, halo and 14-component feature function

The archived original source `prepare_calibration_features_101.py` has now been recovered from the original formation-study `source_snapshot`, not re-created from feature names. Its raw-byte SHA256 is **`1ec1fcb68f947bb6fb76175e249a185d65a094f38060b542df6c3a453c09d90c`**, exactly the value frozen in the original formation reconstruction receipt. The full original file is retained separately as source material. Its top-level code performs historic data-loading actions; for the four-state verification, only the AST-isolated original `metrics` function was evaluated, with no invocation of the file's top-level workflow.

Block selector. On the present periodic `L = 64` lattice, `ell = 8` and `nb = L/ell = 8`. The source takes its frozen selected-index record `sel` and computes `block = int(sel[1]) - (L//4)**2 = int(sel[1]) - 256`. This integer is required to be in `0,...,63`. The selected 8×8 block's upper-left grid indices are `i = ell` $\times$ `floor(block/nb)` and `j = ell` $\times$ `(block mod nb)`; its sites are the 8×8 nonwrapping slice `B = [i,...,i+7]` $\times$ `[j,...,j+7]`. The *external* halo and flux use periodic index wrapping. The selected index is frozen for each formation path and shared across its four ages, not chosen from future outcomes.

Internal block measurements. Let `B_q` and `B_y` be the 64 values of q and y at the selected block. Put `qbar = mean(B_q)`, `ybar = mean(B_y)`, `sigma_{r,c}=1{B_q(r,c) >= 0}`, and define the two directional internal sign-disagreement fractions

$$W_v = \frac{1}{7\times 8}\sum_{r=0}^{6}\sum_{c=0}^{7}\mathbf{1}\{\sigma_{r+1,c}\neq\sigma_{r,c}\}, \qquad W_h = \frac{1}{8\times 7}\sum_{r=0}^{7}\sum_{c=0}^{6}\mathbf{1}\{\sigma_{r,c+1}\neq\sigma_{r,c}\}.$$

Here `var` denotes the *population* variance with divisor the number of included sites (`numpy.var(ddof=0)`), not a sample-variance correction. The wall fraction is `(W_v+W_h)/2` and anisotropy is the **signed** difference `W_v-W_h`.

Periodic halo. Set `R = (i-4,...,i+11) mod L` and `C = (j-4,...,j+11) mod L`. From the 16×16 Cartesian product `R` $\times$ `C`, remove its *central* 8×8 sites with local coordinates `[4:12,4:12]`. The retained **192 cells** form the halo `H` used for means, variances, and q-sign prevalence. The shape and periodic wrapping are part of the original definition and do not follow from the word 'halo' alone.

Boundary flux. For any field `f=q` or `f=y`, define the periodic nearest-neighbor discrete Laplacian `Δ_d f(r,c) = f(r-1,c)+f(r+1,c)+f(r,c-1)+f(r,c+1)-4f(r,c)`. The original source computes exactly

$$\Phi_f = \frac{1}{64}\sum_{(r,c)\in B}\Delta_d f(r,c) = \frac{1}{64}\Bigg[\sum_{\text{outward neighbors}} f - \sum_{\text{corresponding inward edge sites}} f\Bigg].$$

The numerator is the sum over 32 outward boundary edges, with the corresponding inside boundary site subtracted for each edge; corner *sites* contribute twice when they border two outward edges. **The normalization is 64 (block area), not boundary length 32.** Positive flux corresponds to larger summed outside values and negative flux to smaller ones under the code's sign convention. The source itself asserts the equality between the boundary expression and the block-mean discrete Laplacian to 2×10^{-14}.

Table B2. Exact ordered S14 feature definitions (frozen `metrics()` output).

Index	Source name	Exact operation	Current E/Z4 input?
0	q_mean	Mean of 64 block q values, qbar	Yes
1	y_mean	Mean of 64 block y values, ybar	Yes
2	q_var	Mean over block of (q-qbar)^2	Yes
3	wall	(W_v+W_h)/2 from sign disagreement q>=0	Yes
4	q_third_central	Mean over block of (q-qbar)^3	No
5	positive_fraction	Fraction of block sites with q>=0	No
6	wall_anisotropy	W_v-W_h (vertical minus horizontal)	No
7	halo4_q_mean	Mean q over the 192-site periodic halo H	No
8	halo4_y_mean	Mean y over H	No
9	halo4_q_var	Population variance of q over H	No
10	halo4_y_var	Population variance of y over H	No
11	q_boundary_flux	mean_B(Δ_d q)	No
12	y_boundary_flux	mean_B(Δ_d y)	No
13	halo4_positive_fraction	Fraction of H sites with q>=0	No

These operations are transcribed from the source-matched feature function and establish the halo footprint and flux orientation. In particular, y_boundary_flux is the geometric block mean of the discrete spatial Laplacian of y; it does not represent a conserved transport current, and the local y Ornstein–Uhlenbeck evolution includes no spatial diffusion term. E and Z4 depend **only** on indices 0–3; indices 4–13 are not used to fit or score the present two frozen predictors. A revised description of S14 alone therefore neither changes the eight scores nor constitutes a new test.

B3.2 Coarse bins and embedded Z4 midpoint refinements

The frozen E map is np.digitize(x_j, cutpoints_j), with six bins for q_mean and y_mean and five bins for q_var and wall. Cutpoint arrays, in coordinate order, are:

Table B3. Frozen coarse-observation thresholds.

Continuous coordinate	Frozen cutpoints
q_mean	-1, -0.5, 0, 0.5, 1
y_mean	-1, -0.5, 0, 0.5, 1

Continuous coordinate	Frozen cutpoints
q_var	0.25, 0.5, 1, 2
wall	0.05, 0.15, 0.30, 0.50

With left index k = 0,...,number_of_cutpoints, the same fourth-stage source code appends one bit to the original E cell for each of the four sensors when it can define a *finite* [lower, upper] interval. For 0 < k < number_of_cutpoints, the interval uses adjacent cutpoints. For the nonnegative variance and wall lowest bin (k = 0 at indices 2 and 3), the lower bound is zero; for the uppermost wall bin, the upper bound is one. Other tails have no extra split and their bit remains zero. The bit equals 1 when the real sensor value is **at least** the arithmetic midpoint and 0 otherwise. Thus Z4 is an eight-coordinate finite code `(E0, E1, E2, E3, bit0, bit1, bit2, bit3)`, still determined by the same four sensors. These rules are exact frozen-code rules, including equality at the midpoint.

B3.3 The future outcome and an important nonclaim

The O16 target is generated with `np.digitize(terminal selected block q_mean, [-1,-0.5,0,0.5,1])`, producing labels 0–5. The target block is the previously selected $\ell = 8$ sensor block. The last recorded step at lag 16, not a time-window average, supplies the label. Defining X4 as the *unquantized first four* S14 entries gives deterministic maps E = b(X4), Z4 = z(X4), hence E can be reconstructed from Z4 and the information-theoretic data-processing relation in B8 applies. Positive I_B(O;Z4|E) lower confidence does not prove I_B(O;S14|X4)>0, continuous-X4 inadequacy, or S14 sufficiency.

B3.4 Source-byte identity and real-checkpoint regression audit

The recovered source was checked by direct SHA256 against the original frozen receipt before any computation. The original `metrics()` function was AST-isolated, then run on the four saved **real** q/y checkpoints for formation parent `0000` at ages `-8, 0, 16, 20` (chunks 13, 15, 19, 20), retrieved separately from the initial original formation store. Each raw NPZ SHA matched the corresponding entry of the original 2,048-checkpoint prelabel receipt. The freshly evaluated 14-vector was compared with the matching parent/age row of immutable `stage8_frozen_4ages_S14_512.npz`; maximum absolute error across all **56 compared coordinates** was $\mathbf{5.551115123125783 \times 10^{-17}}$. The halo mask contained exactly 192 sites, and both boundary-flux coordinates agreed with an independently constructed four-neighbor discrete-Laplacian mean. This targeted four-state regression check does **not** claim that the present author reprocessed all 2,048 source checkpoints or re-audited all 11,776 original formation files; those have separate historical Colab receipts.

Frozen source: `prepare_calibration_features_101.py`; archived checkpoint source: the original formation `NEW_FORMATION_128 parent0000` directory. The original source and regression-audit JSON/code are retained in the companion `Source_And_Formation_Protocol.zip` archive. The definitions concern measurement provenance; they are not new physical or statistical observations.

B4. Historical training bank and disjoint realized evaluation bank

The fixed E/Z4 lookup was trained from an older 256-parent formation cohort, with 16 futures per age/parent (4,096 class outcomes at each evaluated age). This training cohort's outcomes had been inspected during method development, so selections of sensors, ages, quantizers and smoothing arose within an adaptive historical program; they cannot be advertised as independent hypotheses born before all project activity. For the present test, the lookup JSON SHA256

2d802070905aa7de0e8b4c25477ee82eee018a630099080e50dbc94f2e6893b5 and original scientific evaluator SHA256 fcce87874f501b4d335067467446c176f59563a74525a3a59e3449d8c5e0774f were pinned before prospective class labels existed.

B4.1. Formation-bank initialization and pre-outcome observation-block selection

The original formation-only script run_stage8_formation_pilot.py (SHA256 5983fb5e804b7cd8ec75d7d5dd53ed524d4cccf4fdd684d1e92437b97d0c9453) imports the archived static sampler equilibrium_101.sample (SHA256 083eff6c8af9bf73d531d799ee6f61d148a932bb89f9c399b46579abf6195b84) and the driven quench_pilot_101.advance routine (SHA256 043e7d715b2375fec918d3ea8d4dceaa106fa3afc91b6fe0c3e5ced4fef1b912). Each real formation parent starts with a new q configuration generated by the static, nearest-neighbor q-sector Monte Carlo sampler at r=-0.274 (with the corresponding driven control a=0.726 at time -64 after integrating out y). The script specifies burn=3000, sweeps=1000, thin=5, and alternate hot/cold initialization by the parity of the global parent ID; it takes the sampler's final q configuration as the starting field. This finite static history should not be construed as a proof of perfect equilibrium or independent population sampling.

For each starting q, the script creates a pseudorandom y field as y=q+Normal(0,1) independently at every site, conditional on the chosen q. Using the pre-outcome initialization random stream, it then chooses one non-overlapping block uniformly at each length ℓ=4, 8, 16, 32; the four global block indices are frozen in selected_indices before time evolution. The actual E/Z4/O16 task uses only the selected ℓ=8 block, corresponding to 64 possible 8×8 blocks. This construction does **not** use future outcome labels or optimize a block after measuring prospective performance.

The physical driven path runs from model time -64 to +20 in 21 contiguous chunks. Each chunk has 200 increments at Δt=0.02, so the checkpoint grid is -60, -56, ..., +20 and includes all four primary ages -8, 0, 16, 20 (chunks 13, 15, 19, 20). No physical time is assigned to the static sampler's Monte Carlo sweeps. Each chunk uses the existing ramp evaluated at numerical-step midpoints, with whole-field q/y checkpoints saved between chunks.

The first 128 parents use SeedSequence([10120261020, 101, 8008, parent, stage, chunk]); the later 384 use [10120261021, 101, 8009, parent, stage, chunk], using the same original one() implementation with the namespace variable replaced. Both use distinct global parent ID intervals 0–127 and 128–511, respectively, and their saved protocol records report completion without generating future labels. A seed-ID uniqueness audit is recorded, but seed nonoverlap is not a mathematical proof of population independence. The original script, both frozen protocols, underlying executable kernels and targeted original S14 regression receipts are retained in Source_And_Formation_Protocol.zip.

The new evaluation bank consists of 512 separately generated formation-path identities: 128 pilot formation paths and an additional 384-path extension, all sampled before their new O16 future outcomes were generated. At each of four ages, the corresponding saved complete states were restored, resulting in 2,048 age–parent inputs. Each indexed path may appear in several ages, so the four age banks are *related* and must not be counted as independent formation draws. The frozen features array has shape (4,512,14); ages = [-8,0,16,20]. New evaluation labels have shape (4,512,16). The archival evidence supports disjoint seeded cohort construction and stage-wise identities, subject to ordinary pseudorandom stream assumptions; differing integer seeds or file digests by themselves do not constitute a proof of probabilistic independence.

The lookup's add-one prior and primary hierarchical smoothing with weight 64 follow the main-text subsection "Nested sensor codes and frozen probability forecasts" (Section 4.2). The unobserved-Z4-cell fallback to E occurs for 22.07%, 25.00%, 30.66% and 31.64% of evaluation parents at the four ages, respectively (a prelabel property of the frozen predictor and the new features). The number of distinct *occupied test* E/Z4 cells by age is 41/144, 49/158, 66/196 and 69/203, respectively. Those test-support figures are not extra training samples, and the old 144 generated raw batches discussed in B5 are *not* old 256-parent training futures.

B5. Prospectivity, seed provenance and immutable data chain

B5.1 Recorded order of scientific actions

The following UTC timestamps are supported by the existing archived metadata and the independent 24 Oct audit, not by third-party notarization:

Table B4. Original prospective sequencing recorded in the archive (UTC).

UTC time (2026)	Recorded action	Valid scope
21 Oct, 14:52:26	New 512-bank S14/prelabel receipt created	Parent features available, new future labels not yet created
23 Oct, 01:32:11	Frozen source and lookup G1 package created	Before first new raw future data
23 Oct, 03:01:49	First protocol lock written	Ages, label bins, 16 repeats, eight α events and τ recorded
23 Oct, 03:02:07	First prospective raw batch directory appeared	Earliest archived raw-generation period
23 Oct, 10:00:50–10:02:11	Full 2,048 batches and second-pass receipt completed	Original raw data certified complete, labels not yet extracted
23 Oct, 15:00:36–37	Six-class label, original eight-endpoint evaluation and immutable seal created	Inference sealed after raw completion

Do **not** infer from these records that the G1 package predates formation of the new 512 bank; the claim supported here is freezing before generation of the current new future outcomes, conditional on those already formed inputs. The 144 first-generated raw batches are part of the *new* 512-parent evaluation sample (age -8, parents 0000–0143), not the historically examined 256-parent lookup-training set. The remaining 1,904 batches were generated separately without replacing or reusing the first 144 bytes. The raw output is retained in a separate project data store, outside the statistical replication package.

B5.2 File-level identities and audit levels

Table B5. Frozen source and scientific-result file hashes.

Artifact	SHA256 (archived immutable input/output)
S14 (`4,512,14`)	`6b6bb560d13e061595f36a2a15bc1f0b18a689b86528c64ace274fd00ee921412d802070905aa7de0e8b4c25477ee82e`
Training lookup JSON	`ee018a630099080e50dbc94f2e6893b5fcce87874f501b4d335067467446c176`
Original scientific evaluator	`f59563a74525a3a59e3449d8c5e0774f5f712a0da9a0dd4cbfd81958d9c92d82`
Frozen protocol definition	`31118001dcfd9b983e57901ca66df753f237fca707435b1e4878dd441b480ecd`
O16 class labels (`4,512,16`)	`79878b4143b61db308a0b875c8a4af82c00efbaafd6317ea216298e7235f51ed`
Original one-time statistical output	`c42a9cb9bbf46aac164290bc77d010ef`

The original raw archive contains 2,048 parent–age NPZ batches totaling 2,330,107,618 bytes; the 32,768 continuations each have four segment seeds, for 131,072 unique saved segment identifiers. Source formation audit and raw second-pass metadata/hash reports were completed by the

original Colab run; a later independent audit verified the frozen S14/label/lookup statistical inputs and the eight numerical endpoints, but **did not independently rehash every byte of all 2.33 GB**. Separate 12-file and 15-file raw subsample readbacks are documented. Claims about audit completeness always distinguish these levels of evidence. Filename manifests, reference hashes and seed uniqueness support integrity and no accidental reuse; none by itself certifies initial physical equilibrium or a natural-population law.

The underlying archive contains immutable physical runner sources, formation features, outcome labels, statistical results, generation manifests and separate integrity-audit receipts. The numerical replication package includes only the subset needed to reproduce the frozen statistical endpoints; access details for the full physical archive will be documented separately.

B6. All eight frozen primary endpoints and smoothing sensitivity

All values in this section are **quoted from** the one-time sealed scientific evaluator; this supplementary text is not another run of those tests. Set n = 512, m = 16, one-sided α = 0.00625 for each of four ages and two endpoints per age; τ = 64 primary. The witness radius and log-score radius are calculated from the maximum–minimum range across **all six possible outcome classes**, including classes absent from a particular parent's 16 futures.

Table B6. Frozen values, six class alphabet and one-sided ranges.

Age	$\widehat{\mu}$	r_v	I_{lower}	Mean gain	Log radius	Log lower
-8	0.001832070483	0.001278654189	0.000048729377	0.006087429155	0.016080224989	-0.009992795834
0	0.002002031885	0.001377090225	0.000053557699	0.009653403614	0.012515341615	-0.002861938001
16	0.003503293583	0.001577015857	0.000381557080	0.024393719274	0.011310632199	0.013083087075
20	0.003242303966	0.001550829222	0.000298487051	0.035544754472	0.010990697391	0.024554057081

Positive information certificate: 4/4; positive operational forecast lower bound: 2/4; positive endpoints: 6/8. Both nonpass forecast bounds, at ages -8 and 0, are **negative lower limits for positive observed means**, not evidence for zero true forecast gain. The two endpoints at the same age may be statistically dependent; the union bound does not require independence.

Table B7. Descriptive-only sensitivity with alternative smoothing τ = 256.

Age	Mean log(q_Z4/q_E), τ=256 (nats/outcome)	Status
-8	+0.010161617883	No extra inferential α
0	+0.014494868796	No extra inferential α
16	+0.028972391961	No extra inferential α
20	+0.038674456288	No extra inferential α

τ = 256 was not added as a new confirmatory test. All primary lower bounds remain the τ = 64 original. The physical driving τ_Q = 64 is a different quantity.

B7. After-seal descriptive class and prediction diagnostics

A post-seal, read-only diagnostic analysis used the already sealed labels and S14 features and separately reconstructed fixed q_E/q_Z4 probabilities without refitting. Its cross-entropy difference agrees algebraically, not independently, with the mean log-score difference in B6. No additional p-values, one-sided α levels or direct age comparisons were assigned.

Table B8. Observed terminal six-class counts (each row totals 512 × 16 = 8,192).

Age	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Empirical entropy, nats
-8	18	1488	2573	2549	1550	14	1.376184
0	102	2002	1909	1857	2190	132	1.494025
16	1140	1785	1138	1044	1843	1242	1.764819
20	1418	1607	1064	966	1508	1629	1.773020

Table B9. Posthoc fixed-predictor diagnostics.

Age	Class 0+5	CE_E	CE_{Z4}	$Brier_E$	$Brier_{Z4}$	Cell coverage
-8	0.00390625	1.260430	1.254343	0.686336	0.683960	0.779296875
0	0.02856445	1.302914	1.293260	0.673930	0.669907	0.750000000
16	0.29077148	1.375838	1.351444	0.694209	0.687254	0.693359375
20	0.37194824	1.337839	1.302294	0.671924	0.662623	0.683593750

CE is the same natural-log loss as the primary score contrast, not accuracy in percent. The Brier values use Σ_{class}(q(k) - 1_{Y=k})² per outcome. At later ages more extreme outcomes are observed, and the mean entropy changes appreciably. This is an observed change in the future task's marginal distribution, not proof that physical evolution alone caused the change in Z4 advantage: the formation state, driven future interval and age-specific lookup change together. The Z4 training-cell *coverage* decreases across ages and does not trivially explain increasing mean point gain. All these diagnostics are exploratory/descriptive. Their values are transcribed from separate archived descriptive-diagnostic outputs with immutable input fingerprints.

B8. Conditional probability space, information certificate and multiplicity proof

B8.1 Fixed-bank probability law

Fix a single age t. Let B_t = {x_1,...,x_n} denote the **realized, completely saved** q/y/selected-block states, n = 512. Treat B_t and every trained predictor probability as nonrandom for the current coverage assertion. Let J ~ Uniform{1,...,n}; p_i(y) = P(O=y | B_t,J=i) be the *true* six-class outcome distribution induced by the specified discrete stochastic continuation law and its new random stream. Define a finite mixture

$$P_B(J=i, O=y) = \frac{1}{n} p_i(y), \qquad E_i = b(X4_i), \quad Z_i = z(X4_i), \quad E_i = h(Z_i).$$

Here p_i is a genuine conditional law, not the 16-sample histogram of outcomes and not necessarily either frozen lookup prediction. For a fixed E cell e and refined Z cell z, denote mixture conditionals p_e(y)=P_B(O=y|E=e), p_{ez}(y)=P_B(O=y|E=e,Z=z). Then

$$I_B(O; Z \mid E) = \mathbb{E}_B D_{\mathrm{KL}}(p_{EZ}(\cdot) \| p_E(\cdot)) \geq 0.$$

This finite empirical bank is the *population* for the conditional concentration argument. It is not a fresh draw from an unknown equilibrium distribution over all physical states; averaging over new parents would add another source of uncertainty.

B8.2 Frozen witness and deterministic inequality

Let the two frozen probabilistic lookups be q_Z(y|e,z), q_E(y|e), depending only on their prescribed codes and the **separate old training bank**. Put u_i(y)=q_Z(y|E_i,Z_i)-q_E(y|E_i). Within each E

group, subtract its uniform-parent mean to create v_i(y). Every i in the same Z cell then has the same centered v_i. For all e and outcomes y,

$$\mathbb{E}_B[v_J(y) \mid E = e] = 0, \qquad A = \frac{1}{n}\sum_i \sum_{y=0}^{5} v_i(y)^2.$$

(The inner expression is a deterministic vector norm, not an average over y.) Define the true signed witness moment

$$\mu = \frac{1}{n}\sum_i \sum_y v_i(y)p_i(y) = \mathbb{E}_B\langle v_{EZ}, p_{EZ} - p_E\rangle,$$

where the term with p_E vanishes from E-cell centering. Cauchy–Schwarz gives

$$|\mu|^2 \le A\,\mathbb{E}_B\|p_{EZ} - p_E\|_2^2.$$

For probability vectors p and q, finite-alphabet Pinsker and norm comparison yield

$$\|p - q\|_2^2 \le \|p - q\|_1^2 \le 2D_{\mathrm{KL}}(p\|q).$$

All logarithms are natural. Hence

$$\boxed{|\mu|^2 \le 2A\, I_B(O; Z \mid E).}$$

If A = 0, v ≡ 0 and this witness cannot certify a positive lower bound; reporting zero is valid but **does not prove** I_B = 0. This inequality is an analytic fact about a fixed bank and a frozen witness and does not claim a new fundamental information-theoretic theorem.

B8.3 One-sided Hoeffding radius and independent future repetitions

For each saved parent i, the experiment generates m = 16 independent labeled futures O_i1,...,O_im from p_i, conditional on B_t, a prospectively frozen stream-assignment rule and fixed code. Different parents need not share the same p_i. Their equality or shared formation-path ancestry across **ages** is not required in this single-age conditional argument. The frozen empirical moment is

$$\hat{\mu} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{r=1}^{m} v_i(O_{ir}).$$

Because v_i(O_ir) lies in [min_y v_i(y),max_y v_i(y)] and there are m terms per parent, Hoeffding's one-sided bound yields, with probability at least 1-α,

$$\mu \ge \hat{\mu} - r_v, \qquad r_v = \sqrt{\frac{\log(1/\alpha)}{2n^2m}\sum_{i=1}^{n} R_i^2}, \quad R_i = \max_y v_i(y) - \min_y v_i(y).$$

If μ̂-r_v > 0, the deterministic bound in B8.2 implies I_B ≥ (μ̂-r_v)²/(2A); if μ̂-r_v ≤ 0, the valid reported bound is zero. This is the frozen certificate. The appearance of n²m is exact: Σ_{i,r}range² = mΣ_i R_i², whereas the squared mean denominator is n²m². No parent is being counted 16 times as an independent formation state. If future streams are not probabilistically independent as assumed, the concentration calibration is not certified by seed uniqueness alone.

B8.4 Forecast log-score gain and the eight-event union bound

Set g_i(y)=ln(q_Z(y|E_i,Z_i)/q_E(y|E_i)); the hierarchical positive smoothing prior ensures that g_i(y) is finite on all six classes. Its fixed-bank true average is G = n^{-1}Σ_i Σ_y p_i(y)g_i(y). With all six-class ranges R_i^g, Hoeffding yields a separate one-sided coverage event

$$G \geq \hat{G} - r_g, \qquad r_g = \sqrt{\frac{\log(1/\alpha)\sum_i (R_i^g)^2}{2n^2 m}}, \quad \hat{G} = \frac{1}{nm}\sum_{i,r} g_i(O_{ir}).$$

The original plan specifies two lower-bound events at each of four ages, each α = 0.00625. The union bound is P(all eight lower-coverage statements valid | four realized B_t banks and fixed predictors) ≥ 1-8×0.00625 = 0.95, without requiring independence among ages or endpoints. It **does not** turn a nonpass into a zero effect, produce an age contrast, or give a posterior probability that the full physics hypothesis is correct. The 6/8 count merely records six positive lower bounds after that predetermined allocation. In particular, changing τ or ages on these revealed labels and then treating the result as part of the same prespecified family would invalidate that interpretation.

B8.5 Correct S14 data-processing scope

Let X4 denote the unquantized four-sensor vector, included as the first four coordinates of S14. Because E and Z4 are deterministic functions of X4 and E is determined by Z4, the conditional chain rule implies

$$\boxed{I_B(O; S14 \mid E) \;\geq\; I_B(O; X4 \mid E) \;\geq\; I_B(O; Z4 \mid E).}$$

Therefore every **valid positive certificate for I_B(O;Z4|E)** is also a valid lower certificate for continuous X4 and S14 **relative to the same coarse E**. It does *not* establish I_B(O;S14|X4)>0 or I_B(O;full physical state|S14)=0. Nor does the fact that an unquantized 14-vector uniquely identifies the 512 sampled bank entries prove natural-population sufficiency: finite-bank atomic uniqueness can make conditioning tautological. These are different estimands. The comparison baseline throughout is the same coarse E; this clarification does not alter a numerical endpoint.

B9. Reproduction map, audit scope and limitations

B9.1 What can currently be checked

An authorized analyst with access to the preserved private project archive can (i) verify hashes of the frozen lookup, features, labels, source and evaluator; (ii) recompute E/Z4 codes and smoothed probabilities; (iii) compute the centered witness and the two Hoeffding radii; and (iv) compare all results with the one-time frozen JSON. The initial physical sample and 32,768 future generation require substantially greater computing/storage and must follow the saved checkpoint–seed–protocol plan rather than infer them from manuscript tables.

An independent numerical reimplementation dated 24 October 2026 reported a maximum absolute difference of 4.34×10^{-19} for independently recalculated primary endpoints and 1.11×10^{-16} for descriptive diagnostics, when compared with the sealed records. This is **statistics-level input replication**, not a fresh inspection of all 2.33 GB of original physical trajectories. Earlier raw-level Colab and sample readback records are separate evidence, not replaced by this audit. The present manuscript introduces no new confirmatory tests, revised bins, additional simulations or predictor refitting.

B9.2 Supporting data and limits of the available replay

The derived S14 features, 32,768 class labels, frozen lookup and evaluator, source provenance and an offline arithmetic replay are supplied in the `lattice/` directory of the current companion `Evaluation_Replay.zip` (Online Resource 3). The earlier standalone statistical package was named `Statistical_Replay_Data.zip`; this historical name does not denote a second current submission attachment. The current lattice directory reconstructs the eight primary conditional statistical endpoints. The much larger physical raw-trajectory archive (2,048 input–age batches, approximately 2.33 GB) and the original formation checkpoints are privately retained; a full physical rerun is not included in the supplement. Availability of the full physical archive is a distinct, access-controlled arrangement; its release and long-term preservation cannot be inferred from the compact statistical replay. No public DOI or unrestricted public release for those raw batches is asserted.

B9.3 Reproduction file identities and archived source map

- Frozen physical runner: `101_stage9_O16_gated_future_runner.py`, source `quench_pilot_101.py`, and `RUN_PROTOCOL.json`. Original sources are immutable; earlier draft has not altered them.
- Frozen feature metadata: `stage8_frozen_4ages_S14_512.npz` and `S14_512_prelabel_receipt.json` with 2,048 age–parent source hashes.
- Original feature source `prepare_calibration_features_101.py` (raw bytes, SHA `1ec1fcb68f947bb6fb76175e249a185d65a094f38060b542df6c3a453c09d90c`) recovered from prelabel source snapshot, with exact `metrics()` definitions and a separate four-real-state regression receipt; execute only its isolated `metrics` function when reproducing the features, because the full script contains obsolete project paths and file-writing code.
- Frozen code and predictor: `evaluate_next_independent_cohort_frozen_original.py`, `stage8_frozen_E_Z4_lookup.json`, source hash register and candidate future seed bank.
- Raw completion: `STAGE9J_144_RAW_SHA256_MANIFEST.csv`, `STAGE9K_NEW1904_RAW_SHA256_MANIFEST.csv`, `STAGE9K_FEDERATED2048_RAW_SHA256_MANIFEST.csv`, full raw-second-pass receipt and independent 15-sample source verification.
- Terminal label and statistics: `fresh_O16_512_4ages_16reps_labels.npz`, `101_STAGE9L_FROZEN_E_Z4_ORIGINAL_EVALUATOR_RESULT.json`, and immutable evaluation seal.
- Independent analysis: `101_stage9L_independent_readonly_audit_20261023.py`, its numerical audit JSON, post-seal descriptive code, JSON and CSV, and the 24 Oct 2026 independent audit report.
- Source directories and manifest digests are recorded in the research archive; the present numerical replay is supplied separately.

B9.4 Limitations of the numerical evidence

1. **Input reproducibility limit.** Source identity and a four-checkpoint regression test are documented. The raw source with SHA `1ec1fcb6...` has been recovered and the halo/flux rules are now explicit in B3; four real age-specific checkpoints agree with all 14 frozen feature coordinates. The separate original feature source is preserved in the supporting source archive. This targeted regression is not a fresh full 2,048-checkpoint re-audit.
2. **Population sampling:** The finite-bank concentration bounds are complete, but a population-generalized confidence statement would need additional independent formation paths and specified initial distribution/mixing checks. No such statement is included here.
3. **Driven-kernel Δt convergence:** Existing selected controls are retained, but the nonstationary 64×64 kernel has not received a complete discretization-error budget across the reported O16 outcomes. All results are stated for the frozen discrete update.

4. **Scope of originality:** The finite-information bound uses standard inequalities and is a conditional statistical certificate under a particular executable kernel. The manuscript does not claim a new theorem, optimal predictor, continuous-limit scaling law or universal mechanism.

These limitations delimit the conditional claim and the current data-release status. All numerical findings remain the original frozen endpoints; no new significance claim is made.

B9.5. Positioning relative to previous predictive coarse-graining studies

Table B10. Relation to representative methods; this is a positioning table, not a claim of priority over all literature.

Study	Principal earlier contribution	Different task addressed here
Taverniers et al. (2015) [30]	Learned a coarse stochastic dynamical model from Ising/Glauber data, selected complexity by held-out prediction and compared one-step with autonomous errors	Fixed E/Z4 observational resolution; certificate of strictly positive task-specific information plus finite held-out categorical score bounds; no fitted coarse equation
Tchernookov and Nemenman (2013) [31]	Predictive information in a nonequilibrium critical model, with a focus on predictive-information behavior	Fixed-lag local six-class future within four specified, continuously driven windows; no critical scaling claim
Wang and Tiwary (2021) [32]; Wang et al. (2024) [35]	Used learned predictive information bottlenecks for delayed-state representation / Markov model construction	Prespecified quantization of unchanged physical sensors, audited on a disjoint future bank instead of optimized on the new outcomes
Murphy and Bassett (2024) [34]	Learned information-efficient measurements of chaotic systems	Bounded certificate for a fixed nested sensor resolution, rather than learning optimal measurement channels
Christians et al. (2026) [36]	Learned probabilistic coarse-grained equations including finite-history effects and validated dynamical fidelity	Does not learn governing dynamics or test non-Markovian memory; separately tests conditional information availability and held-out fixed-lookup skill

Related work on predictor-constrained or usable information, such as Xu et al. (ICLR, 2020, arXiv:2002.10689), addresses a broader theory of model-restricted information; the present fixed two-lookup contrast is not the optimized V-information of that framework. The method here applies classical probability inequalities to a finite realized population. Its scientific novelty is the auditable, prospective use of two *different* certification endpoints under a driven statistical-mechanical protocol, combined with documentation of the distinct physical windows and lack of a direct age contrast. It does not establish that the true forecast improvement differs significantly across ages or that additional information is absent at any age.

B9.6. Minimal statistical replay without a new physics simulation

A compact reproduction package contains the immutable original E/Z4 evaluator, frozen training lookup, the $4 \times 512 \times 14$ S14 feature bank, the $4 \times 512 \times 16$ six-class outcome labels, the exact archived evaluation-provenance JSON and the sealed result JSON. Executing the original evaluator with the archived input filenames and source provenance reproduces the sealed scientific JSON exactly on this archive. This **numerical replay** neither repeats formation sampling nor recreates all 2,048 physical batch files; source hashes, training history, physical-correctness limits and original data access are documented separately. The compact companion archive is retained in the separate lattice source archive; the raw physical trajectory archive is maintained separately.
The compact statistical replay described above is supplied in Online Resource 3 (`Evaluation_Replay.zip`, lattice directory), including the original frozen evaluator and evaluation arrays. Its numerical replay does not regenerate formation states or physical future trajectories.

References

[1] Wehmeyer, C., & Noé, F. (2018). Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *The Journal of Chemical Physics*, 148, 241703. doi:10.1063/1.5011399.

[2] Creutzig, F., Globerson, A., & Tishby, N. (2009). Past-future information bottleneck in dynamical systems. *Physical Review E*, 79, 041925. doi:10.1103/PhysRevE.79.041925.

[3] Still, S., Crutchfield, J. P., & Ellison, C. J. (2010). Optimal causal inference: Estimating stored information and approximating causal architecture. *Chaos*, 20, 037111. doi:10.1063/1.3489885.

[4] Sachdeva, V., Mora, T., Walczak, A. M., & Palmer, S. E. (2021). Optimal prediction with resource constraints using the information bottleneck. *PLoS Computational Biology*, 17(3), e1008743. doi:10.1371/journal.pcbi.1008743.

[5] Murphy, K. A., & Bassett, D. S. (2024). Machine-Learning Optimized Measurements of Chaotic Dynamical Systems via the Information Bottleneck. *Physical Review Letters*, 132, 197201. doi:10.1103/PhysRevLett.132.197201.

[6] Krause, A., Singh, A., & Guestrin, C. (2008). Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies. *Journal of Machine Learning Research*, 9, 235–284. JMLR article.

[7] Manohar, K., Brunton, B. W., Kutz, J. N., & Brunton, S. L. (2018). Data-Driven Sparse Sensor Placement for Reconstruction: Demonstrating the Benefits of Exploiting Known Patterns. *IEEE Control Systems Magazine*, 38(3), 63–86. doi:10.1109/MCS.2018.2810460.

[8] Graff, J., Medina, A., & Lagor, F. D. (2023). Information-Based Sensor Placement for Data-Driven Estimation of Unsteady Flows. *AIAA Journal*, 61(11), 4864. doi:10.2514/1.J063015. The cited author version is arXiv:2303.12260.

[9] Attia, A., Alexanderian, A., & Saibaba, A. K. (2018). Goal-Oriented Optimal Design of Experiments for Large-Scale Bayesian Linear Inverse Problems. *Inverse Problems*, 34, 095009. doi:10.1088/1361-6420/aad210.

[10] Argyris, C., Papadimitriou, C., Samaey, G., & Lombaert, G. (2021). A unified sampling-based framework for optimal sensor placement considering parameter and prediction inference. *Mechanical Systems and Signal Processing*, 161, 107950. doi:10.1016/j.ymssp.2021.107950.

[11] Wu, K., Chen, P., & Ghattas, O. (2023). An Offline-Online Decomposition Method for Efficient Linear Bayesian Goal-Oriented Optimal Experimental Design: Application to Optimal Sensor Placement. *SIAM Journal on Scientific Computing*, 45(1), B57–B77. doi:10.1137/21M1466542.

[12] Prigogine, I., & Lefever, R. (1968). Symmetry Breaking Instabilities in Dissipative Systems. II. *The Journal of Chemical Physics*, 48, 1695–1700. doi:10.1063/1.1668896.

[13] Kassam, A.-K., & Trefethen, L. N. (2005). Fourth-Order Time-Stepping for Stiff PDEs. *SIAM Journal on Scientific Computing*, 26, 1214–1233. doi:10.1137/S1064827502410633.

[14] Zhang, H., Ayoub, R., & Sundaram, S. (2017). Sensor selection for Kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms. *Automatica*, 78, 202–210. doi:10.1016/j.automatica.2016.12.025.

[15] Chen, W., Sidky, H., & Ferguson, A. L. (2019). Capabilities and limitations of time-lagged autoencoders for slow mode discovery in dynamical systems. *The Journal of Chemical Physics*, 151(6), 064123. doi:10.1063/1.5112048.

[16] Otto, S. E., Macchio, G. R., & Rowley, C. W. (2023). Learning nonlinear projections for reduced-order modeling of dynamical systems using constrained autoencoders. Chaos, 33, 113130. doi:10.1063/5.0169688.

[17] Linot, A. J., & Graham, M. D. (2022). Data-driven reduced-order modeling of spatiotemporal chaos with neural ordinary differential equations. Chaos, 32, 073110. doi:10.1063/5.0069536.

[18] Schmitt, M. S., Koch-Janusz, M., Fruchart, M., Seara, D. S., Rust, M., & Vitelli, V. (2026). Information theory for dimensionality reduction in dynamical systems. Preprint, arXiv:2312.06608v4. https://arxiv.org/abs/2312.06608v4.

[19] Udupa, S., & Fu, J. (2026). Information-Driven Active Perception for k-step Predictive Safety Monitoring. Preprint, arXiv:2603.23450v1. https://arxiv.org/abs/2603.23450v1.

[20] Kozdoba, M., & Mannor, S. (2026). Score the Algebra, Not the Span: Dimension Reduction for Transfer Operator Models of Dynamical Systems. Preprint, arXiv:2608.18918v1. https://arxiv.org/abs/2608.18918v1.

[21] Dai, R. (2026). Internal compression, collective state formation, and event memory. Author preprint, Zenodo. https://doi.org/10.5281/zenodo.23211854.

[22] Dai, R. (2026). Future-Conditioned Representation and Active Sensing in Dynamical Systems. Version 1, Zenodo. https://doi.org/10.5281/zenodo.23599714.

[23] Hohenberg, P. C. and Halperin, B. I. Theory of dynamic critical phenomena. *Rev. Mod. Phys.* **49**, 435–479 (1977). https://doi.org/10.1103/RevModPhys.49.435

[24] Zurek, W. H. Cosmological experiments in superfluid helium? *Nature* **317**, 505–508 (1985). https://doi.org/10.1038/317505a0

[25] Bray, A. J. Theory of phase-ordering kinetics. *Advances in Physics* **43**, 357–459 (1994). https://doi.org/10.1080/00018739400101505

[26] Mori, H. Transport, collective motion, and Brownian motion. *Progress of Theoretical Physics* **33**, 423–455 (1965). https://doi.org/10.1143/PTP.33.423

[27] Crutchfield, J. P. and Young, K. Inferring statistical complexity. *Phys. Rev. Lett.* **63**, 105–108 (1989). https://doi.org/10.1103/PhysRevLett.63.105

[28] Hoeffding, W. Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **58**, 13–30 (1963). https://doi.org/10.1080/01621459.1963.10500830

[29] Cover, T. M. and Thomas, J. A. *Elements of Information Theory*, 2nd ed. Wiley (2006). https://doi.org/10.1002/047174882X

[30] Taverniers, S., Haut, T. S., Barros, K., Alexander, F. J. and Lookman, T. Physics-based statistical learning approach to mesoscopic model selection. *Phys. Rev. E* **92**, 053301 (2015). https://doi.org/10.1103/PhysRevE.92.053301

[31] Tchernookov, M. and Nemenman, I. Predictive information in a nonequilibrium critical model. *J. Stat. Phys.* **153**, 442–459 (2013). https://doi.org/10.1007/s10955-013-0833-6

[32] Wang, D. and Tiwary, P. State predictive information bottleneck. *J. Chem. Phys.* **154**, 134111 (2021). https://doi.org/10.1063/5.0038198

[33] Bou-Rabee, N. and Vanden-Eijnden, E. Pathwise accuracy and ergodicity of Metropolized integrators for SDEs. *Comm. Pure Appl. Math.* **63**, 655–696 (2010). https://doi.org/10.1002/cpa.20306

[34] Murphy, K. A. and Bassett, D. S. Machine-learning optimized measurements of chaotic dynamical systems via the information bottleneck. *Phys. Rev. Lett.* **132**, 197201 (2024). https://doi.org/10.1103/PhysRevLett.132.197201

[35] Wang, D., Qiu, Y., Beyerle, E. R., Huang, X. and Tiwary, P. An information bottleneck approach for Markov model construction. *J. Chem. Theory Comput.* **20**, 5352–5367 (2024). https://doi.org/10.1021/acs.jctc.4c00449

[36] Christians, L. F., Wojnar, A. and Pak, A. J. Probabilistic forecasting for coarse-grained molecular dynamics. *J. Chem. Theory Comput.* **22**, 4231–4246 (2026). https://doi.org/10.1021/acs.jctc.5c02131

[37] Xu, Y., Zhao, S., Song, J., Stewart, R. and Ermon, S. A theory of usable information under computational constraints. *International Conference on Learning Representations* (2020). https://arxiv.org/abs/2002.10689

[38] Wigren, T. and Schoukens, M. Coupled Electric Drives Data Set and Reference Models. Technical Report 2017-024, Department of Information Technology, Uppsala University (2017). `https://www.nonlinearbenchmark.org/benchmarks/coupled-electric-drives`.

[39] Fisher, A., Rudin, C. and Dominici, F. All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *Journal of Machine Learning Research* **20**(177), 1–81 (2019). `https://www.jmlr.org/papers/v20/18-760.html`.