SYMMETRY, DEFECTS, AND DIFFUSION IN CONTINUOUS-MEMORY RECURRENT NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Recurrent networks that store continuous variables must preserve differences between remembered values despite imperfect updates and state noise. We separate three problems: identifying memory directions, bounding their deterministic distortion, and controlling noise in the decoded coordinates. For ideal dynamics, exact equivariance for each input and uniform nondegeneracy along the trajectory prevent exponential growth or decay in the directions generated by continuous symmetry. For perturbed dynamics, local and finite-horizon bounds distinguish direct memory damage from leakage out of and back into its tangent space. With a fixed transverse contraction gap and no direct damage, the leakage contribution is second order. For noisy dynamics, a decoder-aware local covariance bound yields geometry design rules under a fixed coordinate metric and representation budget: equally scaled orthogonal directions are optimal for isotropic noise with least-squares decoding, whereas unequal noise can favor unequal allocation. Across twenty paired training runs, covariance regularization improves noisy two-angle memory while meeting a specified clean-error equivalence margin; direct noise training incurs a clean-error tradeoff. In coupled four- and eight-angle integrators, the preference between anisotropic and isotropic regularization reverses when evaluation noise changes. These results connect symmetry-protected memory to measurable perturbation bounds and noise-specific representation design.

1 INTRODUCTION

An RNN that remembers heading, position, or phase must preserve distinctions among a continuum of states over long horizons (Chaudhuri & Fiete, 2016). Continuous-attractor models make changes in the stored variable neutral while stabilizing changes away from the memory manifold (Seung, 1996; 1998; Burak & Fiete, 2009); trained navigation networks can also learn path integration and grid-like representations (Cueva & Wei, 2018; Banino et al., 2018; Sorscher et al., 2023). Yet neutral deterministic dynamics is only part of the problem. Model imperfections can pin a memory to a preferred value or make it drift, and state noise can produce wandering even when the ideal dynamics is neutral (Burak & Fiete, 2012; Kilpatrick & Ermentrout, 2013; Agmon & Burak, 2023).

Running example: remembering an angle. Represent a heading by $z(\theta) = r(\cos\theta, \sin\theta)$, with $r > 0$. Motion around the circle changes heading; radial motion leaves the local angle readout unchanged. These are *memory-tangent* and *transverse* directions. An ideal shared-input update preserves heading differences; imperfections can change them directly or through transverse coupling. The same state noise causes larger angular error on a smaller circle. We use this example throughout, then generalize to continuous symmetry orbits.

The paper asks three questions:

1. **Ideal memory:** which state directions preserve differences between stored values without exponential forgetting or amplification?
2. **Imperfect memory:** how do direct perturbations and transverse feedback distort those directions?
3. **Noisy memory:** how does the geometry of the representation convert state noise into decoded error?

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

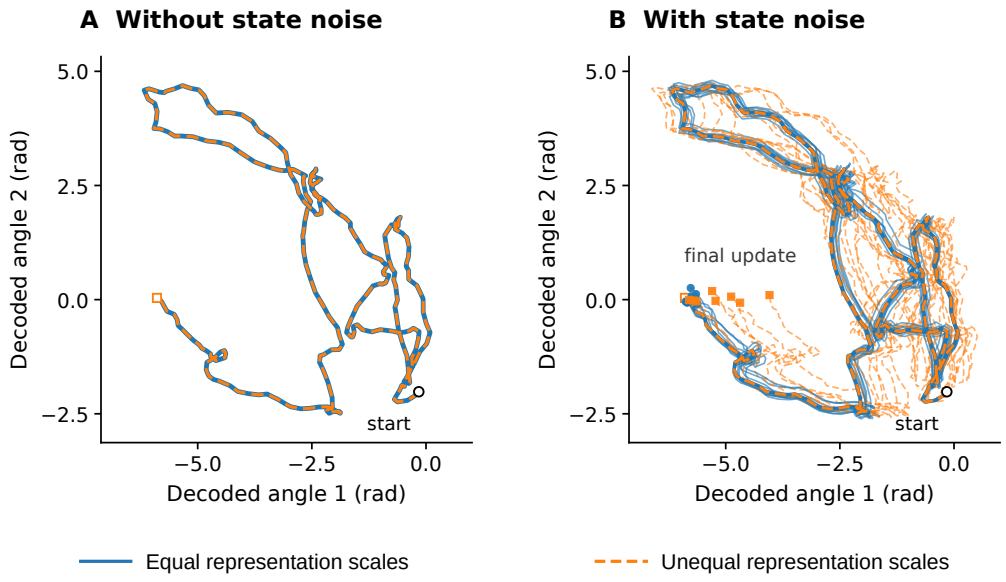


Figure 1: **The same noiseless memory can have different noisy precision.** Two fixed two-angle representations use equal or unequal scales (condition numbers 1 and 8) at the same budget $r_1^2+r_2^2 = 2$. (A) Their noiseless decoded trajectories under a shared input overlap. (B) Eight noisy realizations per representation receive identical isotropic full-state increments within each pair, with covariance $9 \times 10^{-5} I$ per update. Heavy lines show noiseless predictions; lighter lines show individual noisy realizations. Filled circles and squares mark their final update. Axes show unwrapped decoded angles on identical scales. This diagnostic illustrates one preselected checkpoint pair over 256 updates, not eight independent models. The target is omitted and clean task error is nonzero. Selection details and the coordinate precision tradeoff are in Section I.2; multi-seed results and noise-dependent preferences remain in Sections I and K.4.

The common object is the representation derivative: it maps a small memory change to a state displacement. Symmetry identifies it, perturbation blocks describe its distortion, and the decoder maps state changes back to memory. A small Lyapunov exponent alone does not identify this mechanism.

Contributions and scope. We specialize classical equivariant tangent transport (Golubitsky et al., 1988; Krupa, 1990; Rumberger, 2001) to count protected directions, excluding symmetry parameters that leave the state unchanged. Protection alone determines neither defect sensitivity nor decoded noise. Our local and finite-horizon bounds resolve direct damage and transverse feedback, giving more informative bounds than perturbation magnitude alone in the tested comparisons. Our decoder-aware objective predicts when unequal representation scales are preferable; controlled interventions test both distinctions. Generalized least squares and tight-frame inequalities are classical, not separate novelty claims. Selected finite-horizon and architecture-specific results are machine-checked in Lean 4 (de Moura & Ullrich, 2021; The mathlib Community, 2020); Section N.1 specifies coverage and external interfaces.

Experimental setting. Our networks learn angular integration with readouts from circular state blocks. Here T^q denotes q circular coordinates ($S^1 = T^1$). Independent rotations define the symmetry, not uncoupled dynamics: the T^4/T^8 cells share invariant hidden dynamics and cross-block transverse coupling. We test deterministic bounds on S^1/T^2 integrators and noise amplification on prescribed and learned $T^2/T^4/T^8$ geometries. Geometry-only interventions fix nongeometry weights; noise switches test whether geometry preferences change. These controlled tests isolate mechanisms, not generic superiority over unconstrained RNNs. Figure 1 motivates the noise comparison; Table 1 maps the analyses in the appendix.

2 IDEAL CONTINUOUS MEMORY: PROTECTED DIMENSIONALITY

Consider two states differing only by a small heading rotation, driven by the same inputs. If every update respects rotations, the evolved states retain that relationship. The theorem below makes this statement local: the state displacement generated by a small memory change is transported exactly. Uniform nondegeneracy then prevents its exponential collapse or amplification. This preserves differences; it does not by itself ensure that the network integrates the intended velocity correctly.

Consider a controlled recurrent update

$$x_{t+1} = F_{u_t}(x_t), \qquad x_t \in M, \tag{1}$$

where a Lie group G acts smoothly on the finite-dimensional state manifold M. A group element is a transformation, such as a rotation of the circle. The controls here are group invariant: changing the initial heading does not change the supplied velocity. Per-input equivariance means

$$F_u(g\cdot x) = g\cdot F_u(x) \quad \text{for every admissible } u, g, x. \tag{2}$$

Let $\mathfrak{g}$ be the Lie algebra of infinitesimal symmetry parameters. For $\xi \in \mathfrak{g}$, define $A_x\xi = \frac{\mathrm{d}}{\mathrm{d}s}\big|_0 \exp(s\xi)\cdot x$. We also write $\xi_M(x) = A_x\xi$. Thus A_x converts a small symmetry-coordinate change into a state change; its image $E^G_x = T_x(G\cdot x)$ is the memory-tangent space. On the circle, $A_z = (-z_2, z_1)^\top$. In general, some transformations leave the state unchanged: at a reference state they form the *stabilizer* H, with Lie algebra $\mathfrak{h}$. On a fixed orbit-type stratum, the number of independent moving directions is $q = \dim(\mathfrak{g}/\mathfrak{h}) = \dim(G/H)$, not necessarily $\dim G$. For these orbit-type statements, assume a compact group or, more generally, a proper smooth action; stabilizers on one stratum are conjugate, not necessarily the same embedded subgroup (Palais, 1961). Protected dimension is not a general information-capacity or parameter-count statement.

Theorem 1 (Exact tangent transport for controlled recurrent maps)**.** *Let every F_u be C^1, and let x_t be a trajectory of Equation* (1) *contained in a compact set K. Assume Equation* (2)*, $\operatorname{rank} A_x = q$ on K, and that the nonzero singular values of A_x lie in $[c, C]$ with $0 < c \le C < \infty$. If K lies in an orbit-type stratum with representative stabilizer H, then $q = \dim(G/H)$. For every $m > n$,*

$$\begin{aligned} &DF_{u_{m-1}}(x_{m-1})\cdots DF_{u_n}(x_n) \\ &\qquad\cdot A_{x_n} = A_{x_m}. \end{aligned} \tag{3}$$

Consequently, for every $v \in E^G_{x_n}$,

$$\frac{c}{C}\|v\| \le \|P_{m:n}v\| \le \frac{C}{c}\|v\|,$$

where $P_{m:n}$ denotes the displayed tangent product. All q restricted singular-value growth rates are therefore zero, and every nonzero protected vector has zero direct logarithmic growth rate whenever the corresponding limit is defined. For a prescribed nonstationary input, the conclusion is this exact pathwise bound; no stationary spectrum is asserted without an additional driving measure.

Why it holds. Differentiating $F_u(\exp(s\xi)\cdot x) = \exp(s\xi)\cdot F_u(x)$ at $s = 0$ gives $DF_u(x)A_x\xi = A_{F_u(x)}\xi$; iteration gives Equation (3). Equivariance makes the kernels of successive A_{x_t} nested, and constant rank makes them equal. The same nonredundant generator can therefore be used at both endpoints. The bounds on A_x convert exact transport of that generator into the two-sided state-norm bound. The ratio C/c allows bounded distortion, not necessarily exact preservation of Euclidean distance. Section A gives the complete proof.

From transport to Lyapunov exponents. The pathwise result does not require stationary inputs. The stronger statement that the *full* forward spectrum contains at least q zero exponents uses the stationary, ergodic, measurability, integrability, and invertible-base conditions of Corollary 2, stated with its proof in Section A.2. Additional zero exponents are allowed.

For an autonomous equivariant flow, the same differentiation yields

$$\phi_t(\exp(s\xi)\cdot x) = \exp(s\xi)\cdot\phi_t(x) \quad\Longrightarrow\quad D\phi_t(x)\xi_M(x) = \xi_M(\phi_t(x)). \tag{5}$$

This recovers the classical flow identity (Golubitsky et al., 1988; Krupa, 1990). These directions need not coincide with the ordinary time-translation zero mode: the vector field vanishes on a fixed-point orbit, and product groups can give $q > 1$ independent directions. Finite groups have no nonzero infinitesimal generators and do not supply this continuous-memory mechanism (Section A).

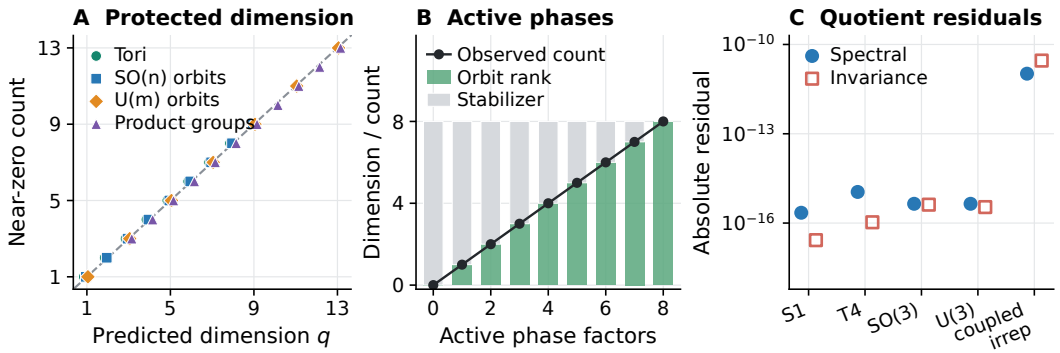


Figure 2: **Ideal-memory orbit geometry.** (A) Across product tori, $SO(n)/SO(n-1)$, $U(m)/U(m-1)$, and product groups, numerical near-zero counts match $q = \dim(G/H)$. Counts support the primary analytical tangent-transport check. (B) Orbit rank and stabilizer dimension stack to eight in the stratified T^8 control; black points show the observed count as phase factors become active. (C) Spectral-match and invariance residuals of the local orbit/quotient Jacobian decomposition. The finite-differenced coupled-irrep Jacobian explains its larger, still $\sim 10^{-11}$ residual.

Does the predicted dimension appear in the models? Figure 2 tests protected dimension, changes in the stabilizer, and removal of the orbit directions from a local Jacobian. Direct analytical tangent transport is the primary diagnostic; thresholded near-zero counts are supporting evidence. Full and reduced Jacobians are compared at the same orbit point. Their largest spectral discrepancy is 1.0×10^{-11} in the finite-differenced coupled-irrep case and at most 1.1×10^{-15} in analytic-Jacobian families. Section B gives the tangent, alignment, and flow-zero controls.

3 Imperfect Continuous Memory: Perturbation Bounds

In the circle example, an imperfection can change the remembered angle directly. It can also send a tangent displacement into radial or other transverse coordinates and then feed it back into the angle. The second route needs both an outward and a return coupling. When direct damage is absent and the transverse contraction gap stays positive, their product sets the second-order leakage term in the bound below. Transverse errors are therefore not automatically harmless.

Four routes through the local dynamics. Select a time map of duration τ at a fixed-point orbit, or make an explicit co-moving identification in which the group-tangent return map is identity. In suitably chosen tangent/normal coordinates the exact map is

$$T_0 = \begin{pmatrix} I & 0 \\ 0 & B \end{pmatrix}, \qquad \|B\| \le \rho < 1. \tag{6}$$

Here E^G denotes the memory tangent and E^N a chosen transverse complement. The latter need not be the naive Euclidean normal space. Tangent invariance alone gives a zero lower-left block; a shear removes any exact upper-right block when $I - B$ is invertible. Its conditioning $\kappa_S = \|S\|\|S^{-1}\|$ must enter ambient-space bounds (Section C.1). Write the perturbed map as

$$T = T_0 + \begin{pmatrix} E_{GG} & E_{GN} \\ E_{NG} & E_{NN} \end{pmatrix}, \quad (a, b, c, e) = \big(\|E_{GG}\|, \|E_{GN}\|, \|E_{NG}\|, \|E_{NN}\|\big). \tag{7}$$

The first index denotes the destination: a is direct memory damage, c sends memory into transverse coordinates, b sends it back, and e changes transverse dynamics. The margin $d = 1 - \rho$ measures transverse contraction; set $D = d - a - e$.

Tracking the memory subspace. A perturbed memory subspace can be written as a *graph* $\{(v, Lv) : v \in E^G\}$: its transverse component is determined by its memory component. This is a subspace graph, not a connectivity graph. The bound $\|L\| \le r_*$ controls its tilt. These graph radii, including r_t below, are distinct from the representation radii r_i used for circular state blocks. The following criterion bounds both tilt and growth.

Proposition 3 (Blockwise invariant-graph criterion and rate bounds). *Suppose $D > 0$, $D^2 \geq 4bc$, and define*

$$r_* = \begin{cases} \dfrac{2c}{D + \sqrt{D^2 - 4bc}}, & b > 0, \\ c/D, & b = 0. \end{cases} \tag{8}$$

If $1 - a - br_* > 0$, the graph transform maps the closed ball $\{L : E^G \to E^N : \|L\| \leq r_*\}$ into itself and hence admits an invariant graph in that ball. On any such graph, set $s = a + br_*$. If $s < 1$, its local memory exponents under repeated iteration of this same selected linear map satisfy

$$\frac{1}{\tau}\log(1 - s) \leq \lambda_G \leq \frac{1}{\tau}\log(1 + s), \tag{9}$$

and its largest ambient principal angle from the exact tangent space obeys

$$\sin\alpha \leq \min\{1, \kappa_S r_*\}. \tag{10}$$

The quantity $s = a + br_*$ is the effective center defect: direct damage plus feedback from the tilted subspace. The graph radius is controlled by c and the contraction margin, so the return term contains both couplings. The criterion is sufficient; failure of its conditions is not proof of instability. It concerns one repeated linear map, not global nonlinear persistence (Fenichel, 1971; Hirsch et al., 1977).

From one map to an input-driven sequence. A network processing inputs encounters different local maps over time. Proposition 4 propagates the tilt and growth bounds through a supplied sequence. Write $J_L v = (v, Lv)$ for the graph embedding; M_t is the induced map on its memory coordinates. Norm and conorm mean its largest and smallest singular values. The coherent frame construction uses a terminal shear and is *offline*: it depends on the future reference sequence, not only past inputs.

Proposition 4 (Finite-horizon transported-graph bounds). *Fix $n < m$. Along a prescribed trajectory, let one coherent frame sequence put the exact reference maps in the form $T_t^0 = \mathrm{diag}(I, B_t)$, $\|B_t\| \leq \rho_t < 1$, and write the perturbed maps as $T_t = T_t^0 + E_t$, with the four (GG, GN, NG, NN) block norms of E_t bounded by (a_t, b_t, c_t, e_t). Starting from a specified graph L_n with $\|L_n\| \leq r_n$, define, for $t = n, \ldots, m-1$,*

$$s_t = a_t + b_t r_t, \qquad r_{t+1} = \frac{c_t + (\rho_t + e_t)r_t}{1 - a_t - b_t r_t}. \tag{11}$$

If every denominator is positive, the specified graph has a unique forward image at each step, and its norm is bounded by the displayed scalar recursion: it remains a graph with $\|L_t\| \leq r_t$ and satisfies $T_t J_{L_t} = J_{L_{t+1}} M_t$. Its center product lies between $\prod_t(1 - s_t)$ and $\prod_t(1 + s_t)$ in conorm and norm. Ambient bounds add only endpoint graph-chart factors, while the largest ambient principal angle α_t from the transported graph to the exact center plane obeys $\sin\alpha_t \leq \min\{1, \kappa(S_t)r_t\}$. For positive step durations τ_t, taking logarithms of the product bounds, including the endpoint factors, and dividing by $\sum_{t=n}^{m-1}\tau_t$ gives finite-horizon physical-time rate intervals. This is finite horizon only: it gives no asymptotic Lyapunov spectrum or bi-infinite invariant bundle.

The numerator in Equation (11) combines new outward leakage with transverse evolution of the existing tilt; the denominator accounts for possible shrinkage of its memory component. Only the endpoint frame factors remain in ambient bounds because intermediate coordinate changes cancel. The complete construction and proof are in Section D. Uniqueness refers to transport of the supplied L_n, not selection of a preferred invariant graph. The constant-map case recovers Proposition 3 when initialized at an invariant graph supplied by that proposition.

Why the route changes perturbative order. With $d > 0$ fixed as $\epsilon \to 0$, direct damage $a = O(\epsilon)$ gives $s = O(\epsilon)$. With $a = e = 0$ and two-way leakage $b, c = O(\epsilon)$,

$$r_* = \frac{c}{d} + O(\epsilon^3/d^3), \qquad s = \frac{bc}{d} + O(\epsilon^4/d^3), \tag{12}$$

The rate-displacement bound is therefore second order and amplified by a smaller contraction gap. One-way leakage alone does not move the center exponent in this constant-block model. These are conditional perturbative bounds, not a universal nonzero scaling coefficient. The derivation and local residual-branch realization are in Section C.

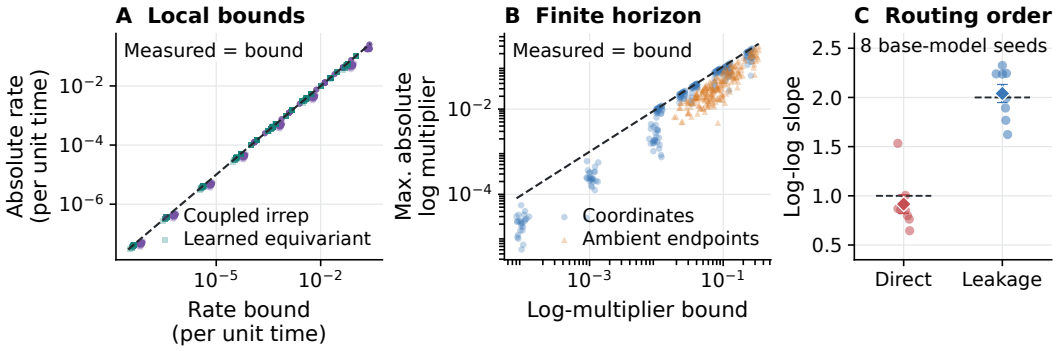


Figure 3: **Local and finite-horizon perturbation bounds.** (A) Absolute physical-time rates versus repeated-map bounds. (B) Maximum absolute log singular values versus the corresponding maximum absolute log bound endpoints, for all 216 valid finite-horizon cases. In A/B, points below the dashed equality line respect the upper bound; distance below indicates conservatism. (C) Eight base-model seeds, with mean $\pm$ SEM: direct 0.916 ± 0.095, leakage 2.040 ± 0.092; dashed references are orders one and two. All 24 assumption-violation controls are retained. The finite-horizon analysis is offline and trajectory conditioned, not an asymptotic or bi-infinite result.

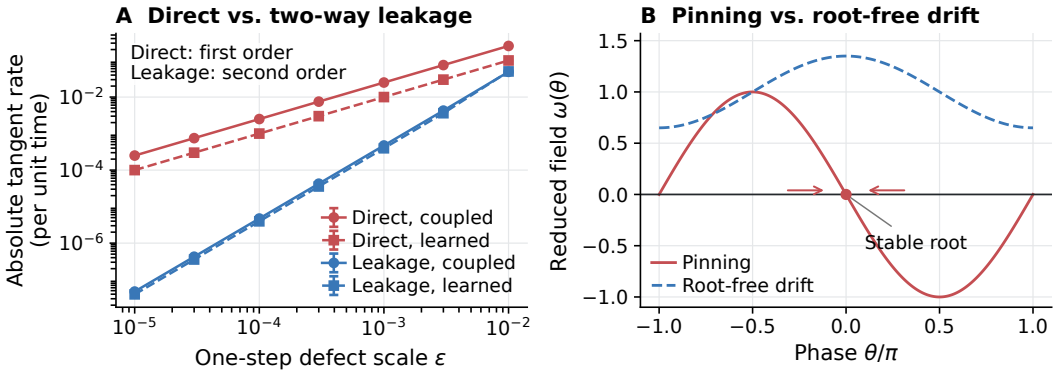


Figure 4: **How defects reach continuous memory.** (A) Slopes fitted within recurrent instances, then summarized over six models per family: direct 1.0006; leakage 2.0038 ± 0.0001 (coupled) and 2.0231 ± 0.0007 (learned), mean $\pm$ SEM. Rates are per physical-time unit. The tiny direct SEM follows from the shared analytic intervention, not universal slope concentration. (B) Illustrative analytic fields $\omega(\theta) = \dot{\theta}/\epsilon$: the marked stable root pins memory; a root-free field instead causes drift with zero asymptotic tangent exponent. Arrows show local motion for $\epsilon > 0$. Thus pseudo-gap opening is not the only symmetry-breaking outcome.

Do the measured perturbations follow the bounds? Across six coupled-irrep instances and six learned S^1 checkpoints, 1,572 of 1,596 nonzero interventions satisfy the conditions and 1,476 give informative bounds, with no measured violations at the declared tolerances. The finite-horizon study covers 216 in-regime cases and retains 24 out-of-regime controls (Figure 3). Native residuals and nonlinear pinning-time checks appear in Sections F and G, not as global error guarantees.

Does resolving the blocks improve the bound? A retrospective comparator uses the same frames, contraction information, and procedure, but sets $a = b = c = e = \delta = \|E\|_2$. Block resolution gives valid bounds in 2,515 of 2,520 ambient-residual cases, versus 2,160 for this norm-only comparator. Both methods cover 216 of 240 finite-horizon sequences; block resolution gives narrower intervals on the common valid subset (Section K.3). These nested cases are not independent models, and the comparison does not cover every possible norm-only bound.

A zero rate need not mean an accurate memory. Figure 4 returns to the heading example. For $\dot{\theta} = \epsilon\omega(\theta)$, $\epsilon > 0$, a stable root pins the memory and has $\lambda_{\mathrm{sym}} = \epsilon\omega'(\theta_*) < 0$; the positive magnitude $\gamma = -\lambda_{\mathrm{sym}}$ is a *pseudo-gap*. But a continuous periodic field with no root has fixed sign and magnitude bounded away from zero. The phase drifts by $O(\epsilon t)$ while the tangent multiplier $\omega(\theta_t)/\omega(\theta_0)$ stays bounded, giving zero asymptotic exponent. Thus a rate change detects contraction or expansion, not every systematic memory error (Kilpatrick & Ermentrout, 2013; Agmon & Burak, 2023).

4 Noisy continuous memory: local precision

Neutral deterministic dynamics does not prevent noise-driven wandering (Burak & Fiete, 2012; Kilpatrick & Ermentrout, 2013). Return to $z = r(\cos\theta, \sin\theta)$. The local angle decoder gives

$$\delta\theta \approx \frac{(-\sin\theta, \cos\theta)\,\delta z}{r}, \qquad D_\theta = \frac{\sigma^2}{r^2} \quad \text{when the state covariance rate is } \sigma^2 I.$$

A smaller circle therefore amplifies the same state noise into a larger angular error. With several stored coordinates and a fixed total scale, enlarging one circle leaves less for the others. The general result below expresses this tradeoff for a specified representation, noise, and decoder.

From state noise to memory noise. Remove symmetry parameters that do not move the state to obtain q independent memory coordinates. Fix their local quotient-coordinate metric and choose an isometric horizontal frame $\iota_x : \mathbb{R}^q \to (\ker A_x)^\perp \subset \mathfrak{g}$. The reduced action

$$\bar{A}_x := A_x \iota_x : \mathbb{R}^q \longrightarrow E_x^G. \tag{13}$$

maps a small memory-coordinate change to a state displacement. Its singular values measure representation sensitivity in that chosen metric; its condition number compares the strongest and weakest directions. In the free torus experiments, the natural coordinates give $\bar{A}_x = A_x$.

Conversely, the decoder derivative $Z_x = D\psi_x$ maps a state displacement to a change in the decoded coordinates. Correct decoding to first order along the orbit means that the two maps compose to identity:

$$Z_x \bar{A}_x = I_q. \tag{14}$$

For an ambient Itô increment with covariance $Q_x\,\mathrm{d}t$, Itô's formula (Øksendal, 2003) gives

$$D_\psi(x) = Z_x Q_x Z_x^\top. \tag{15}$$

Thus representation geometry alone is insufficient: a decoder may also respond to transverse noise. Here D_ψ is a covariance-rate matrix; the Fokker–Planck convention places a factor $1/2$ before it. Curvature can produce Itô drift, and long-time diffusion also depends on the surrounding dynamics and isochrons (MacLaurin, 2023; Ahsan et al., 2025).

For a discrete update followed by noise of covariance $Q_y \Delta t$, the leading decoded covariance is $Z_y Q_y Z_y^\top \Delta t$ at the updated state $y = F(x)$. Corollary 5 in Section H.2 states the required remainder assumption and the $o(\Delta t)$ error. This local small-step bridge does not assert an exact nonlinear finite-horizon identity.

What is the best faithful decoder?

Theorem 6 (Decoder-aware isotropic noise floor)**.** *Assume $Q_x = \sigma^2 I$ with $\sigma > 0$. Every local decoder satisfying Equation* (14) *obeys*

$$D_\psi(x) \succeq \sigma^2 (\bar{A}_x^\top \bar{A}_x)^{-1}. \tag{17}$$

Equality holds if and only if $Z_x = \bar{A}_x^\dagger$, equivalently the decoder has no first-order sensitivity in directions orthogonal to $\mathrm{range}(\bar{A}_x) = E_x^G$.

Indeed, $Z = \bar{A}^\dagger + R$ with $R\bar{A} = 0$ gives $ZZ^\top = (\bar{A}^\top \bar{A})^{-1} + RR^\top$. Extra sensitivity away from the orbit adds covariance. For structured positive-definite noise, the minimum-covariance left inverse is

$$Z_* = (\bar{A}^\top Q^{-1} \bar{A})^{-1} \bar{A}^\top Q^{-1}, \qquad D_{\psi,*} = (\bar{A}^\top Q^{-1} \bar{A})^{-1}. \tag{18}$$

This is the classical generalized least-squares solution (Aitken, 1936), interpreted here as a local recurrent-memory noise floor.

How should a fixed representation budget be allocated? For circular blocks, $P = \mathrm{tr}(\bar{A}_x^\top \bar{A}_x) = \sum_i r_i^2$ is the sum of squared radii. More generally it is the squared sensitivity summed over memory-coordinate directions. The next result prevents reducing noise merely by rescaling every direction without cost.

Corollary 7 (Scaled-isometric action optimum)**.** *Under the assumptions of Theorem 6, fix tangent energy $P = \mathrm{tr}(\bar{A}_x^\top \bar{A}_x)$. Then every faithful decoder satisfies*

$$\lambda_{\max}(D_\psi) \geq \frac{\sigma^2 q}{P}, \qquad \mathrm{tr}(D_\psi) \geq \frac{\sigma^2 q^2}{P}. \tag{19}$$

Both lower bounds are attained exactly when $\bar{A}_x^\top \bar{A}_x = (P/q) I_q$ and $Z_x = \bar{A}_x^\dagger$.

Under isotropic noise, equal sensitivity in orthogonal coordinate directions is optimal with least-squares decoding: the Gram matrix is proportional to identity. This is the scaled-isometry or tight-frame condition (Benedetto & Fickus, 2003; Xu et al., 2023; 2025). The conclusion depends on the coordinate metric, noise model, decoder, and budget; it is not a universal preference for Euclidean isotropy.

Does geometry predict measured noise amplification? We fix $P = 2$ and prescribe $\kappa(\bar{A}) \in \{1, 1.5, 2, 3, 5, 8\}$ in T^2 integrators. Radius rescaling leaves the noiseless input–output behavior matched by construction. Six seed identities are paired across geometries, not 36 independent models, and the angle decoder equals $\bar{A}^\dagger$ on the target orbit. Median covariance-tensor error is 0.56% locally and 3.32% over long zero-input simulations. Increasing κ from 1 to 8 raises noisy horizon-256 worst-coordinate RMSE by 0.639 ± 0.011 rad (mean $\pm$ SEM), with all six pairs increasing. Gaussian increments follow updates at $\Delta t = 0.1$; the long-horizon simulations test nonlinear composition beyond the small-step bridge (Section I).

Can learning improve geometry without the same clean-error cost? We train radii $r_i^2 = 2\,\mathrm{softmax}(\alpha)_i$ from $\kappa = 8$ using

$$\mathcal{L}_{\mathrm{train}} = \mathcal{L}_{\mathrm{clean}} + 10^{-3}\mathcal{L}_{\mathrm{geom}}, \qquad \mathcal{L}_{\mathrm{geom}} = \tfrac{1}{2}\,\mathrm{tr}\big[(\bar{A}^\top \bar{A})^{-1}\big] - 1.$$

The penalty shapes radii during task training; symmetry and angle decoding remain fixed. The analytical allocation supplies a local-covariance reference, while the learned training conditions test whether reducing this covariance improves finite-horizon memory. A preliminary clean-only study selected the coefficient without using noisy test error (Section J.2). We then evaluated the fixed settings in an independent replication with twenty paired training runs (Section K), without pooling historical observations. Noisy horizon-256 worst-coordinate RMSE improves in $20/20$ pairs: mean change -0.623 rad (95% paired t interval $[-0.630, -0.616]$), with its clean-change interval inside the specified ± 0.01 rad equivalence margin. The isotropic-oracle gap remains 0.050 ± 0.001 rad (mean $\pm$ SEM). Noise training also improves every noisy pair but incurs a clean change of $+0.018$ rad (95% interval $[0.006, 0.031]$), failing equivalence (Figure 5A,B).

Does the preferred geometry change when the noise changes? For coupled T^q cells with angle decoding and block noise $Q_i = \sigma^2 v_i I_2$, $D_\psi = \mathrm{diag}(\sigma^2 v_i / r_i^2)$. At fixed $\sum_i r_i^2 = P$, its trace is minimized by $r_i^2 = P\sqrt{v_i}/\sum_j \sqrt{v_j}$: noisier coordinates receive more squared radius. The variance-weighted penalty is given in Section K.4. This trace optimum is local, not a guarantee of optimal nonlinear task error. Sixteen paired training units per dimension use $P = q$, fixing energy per coordinate. At horizon 256 (time 25.6), speed 1.5, anisotropic-covariance minus isotropic regularization changes mean endpoint phase MSE by -0.006106 rad^2 for T^4 (primary; 95% paired t interval $[-0.006674, -0.005538]$) and -0.006188 for T^8 (secondary). Switching the same cells to isotropic noise reverses the differences to $+0.014229$ and $+0.009814$ rad^2 (Figure 5C). Geometry suited to anisotropic noise is not necessarily best under isotropic noise.

To isolate geometry, we also change only the final clean-trained cells' radius allocation, holding nongeometry weights, inputs, and noise streams fixed. Trace-optimal swaps lower anisotropic-noise MSE in all sixteen units per dimension, with absolute paired mean clean changes at most 2.1×10^{-10} rad^2. Isotropic swaps improve the mean but worsen one T^4 unit (Section K.4). These interventions isolate geometry without attributing all differences between training conditions to it or asserting deterministic contraction.

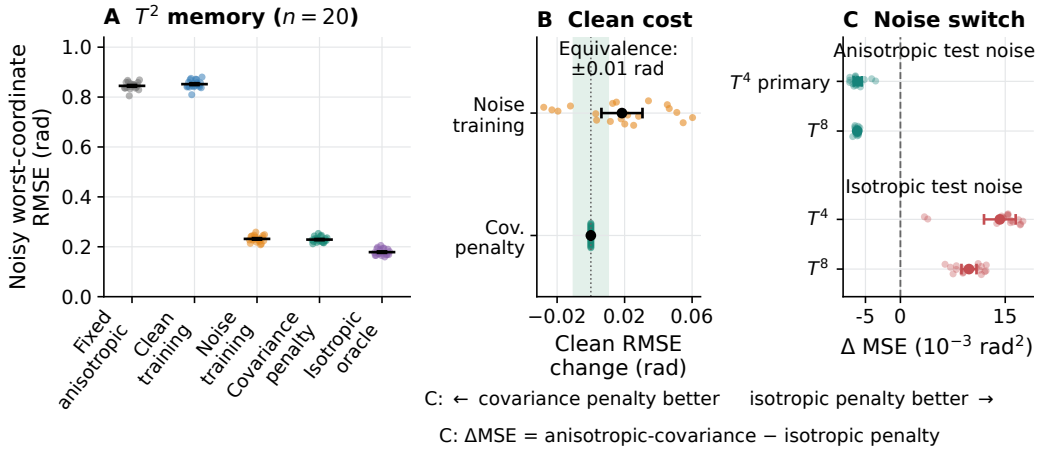


Figure 5: **Geometry improves noisy memory, but its preferred allocation depends on the noise.** (A) Horizon-256 T^2 worst-coordinate RMSE: twenty paired training units; black marks are mean $\pm$ SEM. (B) Clean RMSE changes relative to clean training, with paired 95% t intervals and the shaded ± 0.01 rad equivalence margin; covariance but not noise training meets it. (C) Anisotropic-covariance minus isotropic regularization: sixteen paired units per dimension, with 95% t intervals; the primary row is labeled. The subtraction is unchanged across test-noise conditions. A/B use rad; C ticks are in 10^{-3} rad^2. These are conditional task consequences, not theorem proof or a universally optimal geometry.

5 Related work

Continuous-attractor models and equivariant networks motivate structured memory representations (Seung, 1996; 1998; Burak & Fiete, 2009; Cohen & Welling, 2016; Bronstein et al., 2021). The closest recurrent studies link attractors to translation equivariance, determine symmetry-shaped manifolds, and develop flow-equivariant sequence and world models (Zhang et al., 2022; Di Bernardo et al., 2025; Keller, 2025; Lillemark et al., 2026); robust memory can also arise without exact symmetry (Can & Krishnamurthy, 2025). We connect identifiable memory tangents to local and finite-horizon defect routing and decoder-conditioned noise. The perturbation tools are specializations of classical theory, not a general dichotomy theorem. Relative to Mo (2026), this work adds controlled transport, finite-horizon bounds, and decoder-aware learning. Section L distinguishes control-input group models, symmetry relaxation, Goldstone-like propagation, simulation guarantees, stochastic manifold models, and geometric navigation, and identifies the classical estimation and frame ingredients.

6 Using the framework and limitations

Reliable continuous memory requires more than neutral dynamics. First identify the state directions generated by a known continuous action and check their rank and conditioning. Then decompose measured derivative perturbations into tangent and transverse blocks to evaluate the local criterion or the finite-horizon recursion. Finally, specify the decoder, state-noise covariance, coordinate metric, and budget to predict local decoded noise and choose a geometry objective. This is not a black-box method for discovering an unknown symmetry, and failure of a sufficient bound does not establish instability.

These conclusions have distinct domains. Exact protection requires per-input equivariance and a nondegenerate orbit stratum. The time-varying perturbation bound is offline, assumes normal contraction in its chosen coordinates, and makes no asymptotic, online, or bi-infinite claim. The T^4/T^8 noise tests use a fixed angle decoder and block noise; they establish neither global stochastic stability nor deterministic contraction. Within these limits, the framework explains what preserves memory differences, how defects distort them, and why changing the noise can change the best representation geometry.

REPRODUCIBILITY STATEMENT

Complete proofs and experimental settings are provided in the appendix. The anonymous supplement contains the Lean development, theorem-to-declaration mapping, Python implementation, study configurations, checkpoints, result records, and figure-generation or correction scripts. Numerical precision, random-stream separation, and unsuccessful runs are included with the corresponding studies; Section N summarizes the computational resources and verification scope.

AI USE STATEMENT

Large language models assisted with critiquing mathematical claims, revising proofs, and implementing the Lean formalization. The formal and executable checks, including their limitations, are documented in Section N.1 and the supplementary material; they do not establish end-to-end formal verification of the manuscript or experimental implementation. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Haggai Agmon and Yoram Burak. Simultaneous embedding of multiple attractor manifolds in a recurrent neural network using constrained gradient optimization. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/361e5112d2eca09513bbd266e4b2d2be-Abstract-Conference.html.

Zaid Ahsan, Harry Dankowicz, and Christian Kuehn. Adjoint-based projections for quantifying statistical covariance near stochastically perturbed limit cycles and tori. *SIAM Journal on Applied Dynamical Systems*, 24(2):1455–1493, 2025. doi: 10.1137/23M1599318.

A. C. Aitken. On least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh*, 55:42–48, 1936. doi: 10.1017/S0370164600014346.

Martin Arjovsky, Amar Shah, and Yoshua Bengio. Unitary evolution recurrent neural networks. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 1120–1128, 2016. URL https://proceedings.mlr.press/v48/arjovsky16.html.

Andrea Banino, Caswell Barry, Benigno Uria, Charles Blundell, Timothy Lillicrap, Piotr Mirowski, Alexander Pritzel, Martin J. Chadwick, Thomas Degris, Joseph Modayil, Greg Wayne, Hubert Soyer, Fabio Viola, Brian Zhang, Ross Goroshin, Neil Rabinowitz, Razvan Pascanu, Charles Beattie, Stig Petersen, Amir Sadik, Stephen Gaffney, Helen King, Koray Kavukcuoglu, Demis Hassabis, Raia Hadsell, and Dharshan Kumaran. Vector-based navigation using grid-like representations in artificial agents. *Nature*, 557(7705):429–433, 2018. doi: 10.1038/s41586-018-0102-6.

John J. Benedetto and Matthew Fickus. Finite normalized tight frames. *Advances in Computational Mathematics*, 18:357–385, 2003. doi: 10.1023/A:1021323312367.

G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 2: Numerical application. *Meccanica*, 15(1):21–30, 1980a. doi: 10.1007/BF02128237.

G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 1: Theory. *Meccanica*, 15(1):9–20, 1980b. doi: 10.1007/BF02128236.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021. doi: 10.48550/arXiv.2104.13478. URL https://arxiv.org/abs/2104.13478.

Yoram Burak and Ila R. Fiete. Accurate path integration in continuous attractor network models of grid cells. *PLOS Computational Biology*, 5(2):e1000291, 2009. doi: 10.1371/journal.pcbi.1000291.

Yoram Burak and Ila R. Fiete. Fundamental limits on persistent activity in networks of noisy neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 109 (43):17645–17650, 2012. doi: 10.1073/pnas.1117386109.

Tankut Can and Kamesh Krishnamurthy. Emergence of robust memory manifolds. *PRX Life*, 3(2): 023006, 2025. doi: 10.1103/PRXLife.3.023006.

Rishidev Chaudhuri and Ila Fiete. Computational principles of memory. *Nature Neuroscience*, 19(3): 394–403, 2016. doi: 10.1038/nn.4237.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734, 2014. doi: 10.3115/v1/D14-1179.

Matteo Cipollina, Michail Karatarakis, and Freek Wiedijk. Formalized hopfield networks and boltzmann machines, 2025.

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

W. A. Coppel. *Dichotomies in Stability Theory*, volume 629 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1978. doi: 10.1007/BFb0067780.

Madison Cotteret, Christopher J. Kymn, Hugh Greatorex, Martin Ziegler, Elisabetta Chicca, and Friedrich T. Sommer. High-resolution spatial memory requires grid-cell-like neural codes. *arXiv preprint arXiv:2507.00598*, 2025.

Christopher J. Cueva and Xue-Xin Wei. Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=B17JTOe0-.

Ran Darshan and Alexander Rivkind. Learning to represent continuous variables in heterogeneous neural networks. *Cell Reports*, 39(1):110612, 2022. doi: 10.1016/j.celrep.2022.110612.

Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970. doi: 10.1137/0707001.

Leonardo de Moura and Sebastian Ullrich. The Lean 4 theorem prover and programming language. In *Automated Deduction – CADE 28*, volume 12699 of *Lecture Notes in Computer Science*, pp. 625–635. Springer, 2021. doi: 10.1007/978-3-030-79876-5_37.

Arianna Di Bernardo, Adrian Valente, Francesca Mastrogiuseppe, and Srdjan Ostojic. Shaping manifolds in equivariant recurrent neural networks. *arXiv preprint arXiv:2511.04802*, 2025. doi: 10.48550/arXiv.2511.04802. URL https://arxiv.org/abs/2511.04802.

Rainer Engelken. Gradient flossing: Improving gradient descent through dynamic control of jacobians. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/214ce905bf2072535e34b3cf873cbbc8-Abstract-Conference.html.

Neil Fenichel. Persistence and smoothness of invariant manifolds for flows. *Indiana University Mathematics Journal*, 21(3):193–226, 1971. doi: 10.1512/iumj.1972.21.21017.

Gary Froyland, Simon Lloyd, and Anthony Quas. A semi-invertible Oseledets theorem with applications to transfer operator cocycles. *Discrete and Continuous Dynamical Systems*, 33(9):3835–3860, 2013. doi: 10.3934/dcds.2013.33.3835.

Robert Joseph George, Jennifer Cruden, Will Adkisson, Xiangru Zhong, Huan Zhang, and Anima Anandkumar. TorchLean: Formalizing neural networks in Lean, 2026.

F. Ginelli, P. Poggi, A. Turchi, H. Chaté, R. Livi, and A. Politi. Characterizing dynamics with covariant lyapunov vectors. *Physical Review Letters*, 99(13):130601, 2007. doi: 10.1103/PhysRevLett.99.130601.

Martin Golubitsky, Ian Stewart, and David G. Schaeffer. *Singularities and Groups in Bifurcation Theory, Volume II*, volume 69 of *Applied Mathematical Sciences*. Springer, 1988. doi: 10.1007/978-1-4612-4574-2.

Mikael Henaff, Arthur Szlam, and Yann LeCun. Recurrent orthogonal networks and long-memory tasks. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2034–2042, 2016. URL https://proceedings.mlr.press/v48/henaff16.html.

Morris W. Hirsch, Charles C. Pugh, and Michael Shub. *Invariant Manifolds*, volume 583 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1977. doi: 10.1007/BFb0092042.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

Nabil Iqbal, T. Anderson Keller, Yue Song, Takeru Miyato, and Max Welling. Spontaneous symmetry breaking and goldstone modes for deep information propagation. *arXiv preprint arXiv:2605.14685*, 2026. doi: 10.48550/arXiv.2605.14685. URL https://arxiv.org/abs/2605.14685.

T. Anderson Keller. Flow equivariant recurrent neural networks. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.48550/arXiv.2507.14793. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/e637029c42aa593850eeebf46616444d-Abstract-Conference.html. Spotlight.

Zachary P. Kilpatrick and Bard Ermentrout. Wandering bumps in stochastic neural fields. *SIAM Journal on Applied Dynamical Systems*, 12(1):61–94, 2013. doi: 10.1137/120877106.

Andrew V. Knyazev and Merico E. Argentati. Principal angles between subspaces in an A-based scalar product: Algorithms and perturbation estimates. *SIAM Journal on Scientific Computing*, 23 (6):2008–2040, 2002. doi: 10.1137/S1064827500377332.

Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2747–2755, 2018. URL https://proceedings.mlr.press/v80/kondor18a.html.

Martin Krupa. Bifurcations of relative equilibria. *SIAM Journal on Mathematical Analysis*, 21(6): 1453–1486, 1990. doi: 10.1137/0521081.

Arthur Liang, Ábel Ságodi, Piotr A. Sokół, and Il Memming Park. Symmetry-regularized learning of continuous attractor dynamics. In *NeurIPS 2025 Workshop on Symmetry and Geometry in Neural Representations (NeurReps)*, 2025. URL https://openreview.net/forum?id=W8Gf7CYCo8. Poster.

Hansen Jin Lillemark, Benhao Huang, Fangneng Zhan, Yilun Du, and T. Anderson Keller. Flow equivariant world models: Memory for partially observed dynamic environments. In *International Conference on Machine Learning*, 2026. doi: 10.48550/arXiv.2601.01075. URL https://arxiv.org/abs/2601.01075. Accepted.

J. MacLaurin. Phase reduction of waves, patterns, and oscillations subject to spatially extended noise. *SIAM Journal on Applied Mathematics*, 83(3):1215–1244, 2023. doi: 10.1137/21M1451221.

Hanson Hanxuan Mo. Symmetry-protected lyapunov neutral modes in equivariant recurrent networks. *arXiv preprint arXiv:2605.03338*, 2026. doi: 10.48550/arXiv.2605.03338. URL https://arxiv.org/abs/2605.03338.

Ramón Nartallo-Kaluarachchi, Renaud Lambiotte, and Alain Goriely. Drift–diffusion matching: Embedding dynamics in latent manifolds of asymmetric neural networks. *arXiv preprint arXiv:2602.14885*, 2026. doi: 10.48550/arXiv.2602.14885. URL https://arxiv.org/abs/2602.14885.

Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Universitext. Springer Berlin, Heidelberg, 6 edition, 2003. doi: 10.1007/978-3-642-14394-6.

V. I. Oseledets. A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems. *Transactions of the Moscow Mathematical Society*, 19:197–231, 1968.

Richard S. Palais. On the existence of slices for actions of non-compact lie groups. *Annals of Mathematics*, 73(2):295–323, 1961. doi: 10.2307/1970335.

Stefanos Pertigkiozoglou, Mircea Petrache, Shubhendu Trivedi, and Kostas Daniilidis. Recurrent equivariant constraint modulation: Learning per-layer symmetry relaxation from data. *arXiv preprint arXiv:2602.02853*, 2026. doi: 10.48550/arXiv.2602.02853. URL https://arxiv.org/abs/2602.02853.

Christian Pötzsche. *Geometric Theory of Discrete Nonautonomous Dynamical Systems*, volume 2002 of *Lecture Notes in Mathematics*. Springer, Berlin, Heidelberg, 2010. doi: 10.1007/978-3-642-14258-1.

Matthias Rumberger. Lyapunov exponents on the orbit space. *Discrete and Continuous Dynamical Systems*, 7(1):91–113, 2001. doi: 10.3934/dcds.2001.7.91. URL https://www.aimsciences.org/article/doi/10.3934/dcds.2001.7.91.

Robert J. Sacker and George R. Sell. Existence of dichotomies and invariant splittings for linear differential systems, iii. *Journal of Differential Equations*, 22(2):497–522, 1976. doi: 10.1016/0022-0396(76)90043-7.

Ábel Ságodi and Il Memming Park. Universal approximation theorems for dynamical systems with infinite-time horizon guarantees. *arXiv preprint arXiv:2602.08640*, 2026. doi: 10.48550/arXiv.2602.08640. URL https://arxiv.org/abs/2602.08640.

Ábel Ságodi, Guillermo Martín-Sánchez, Piotr Sokół, and Il Memming Park. Back to the continuous attractor. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/7b78a2a7360d5a9ad750834dc5a33bfb-Abstract-Conference.html.

H. S. Seung. How the brain keeps the eyes still. *Proceedings of the National Academy of Sciences of the United States of America*, 93(23):13339–13344, 1996. doi: 10.1073/pnas.93.23.13339.

H. S. Seung. Continuous attractors and oculomotor control. *Neural Networks*, 11(7–8):1253–1258, 1998. doi: 10.1016/S0893-6080(98)00064-1.

Ben Sorscher, Gabriel C. Mel, Samuel A. Ocko, Lisa M. Giocomo, and Surya Ganguli. A unified theory for the computational and mechanistic origins of grid cells. *Neuron*, 111(1):121–137.e13, 2023. doi: 10.1016/j.neuron.2022.10.003.

G. W. Stewart. Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM Review*, 15(4):727–764, 1973. doi: 10.1137/1015095.

The mathlib Community. The Lean mathematical library. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs*, pp. 367–381. ACM, 2020. doi: 10.1145/3372885.3373824.

Margot Wagner, Yusi Chen, Arjun Karuvally, Mia Cameron, and Terrence J. Sejnowski. Balancing stability and flow in hippocampal networks via inductive bias and learned symmetry breaking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 381(1954):20250260, 2026. doi: 10.1098/rstb.2025.0260.

Hongbo Wang. Certified world models: Predictability across configuration, horizon, and resolution. *arXiv preprint arXiv:2606.13092*, 2026a. doi: 10.48550/arXiv.2606.13092. URL https://arxiv.org/abs/2606.13092.

Hongbo Wang. Conformal orbit-valid trust horizons for equivariant world models. *arXiv preprint arXiv:2606.24946*, 2026b. doi: 10.48550/arXiv.2606.24946. URL https://arxiv.org/abs/2606.24946.

Rui Wang, Robin Walters, and Rose Yu. Approximately equivariant networks for imperfectly symmetric dynamics. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 23078–23091. PMLR, 2022. URL https://proceedings.mlr.press/v162/wang22aa.html.

Per-Åke Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12:99–111, 1972. doi: 10.1007/BF01932678.

Dehong Xu, Ruiqi Gao, Wen-Hao Zhang, Xue-Xin Wei, and Ying Nian Wu. Conformal isometry of Lie group representation in recurrent network of grid cells. In *Proceedings of the 1st NeurIPS Workshop on Symmetry and Geometry in Neural Representations*, volume 197 of *Proceedings of Machine Learning Research*, pp. 370–387. PMLR, 2023. URL https://proceedings.mlr.press/v197/xu23a.html.

Dehong Xu, Ruiqi Gao, Wenhao Zhang, Xue-Xin Wei, and Ying Nian Wu. On conformal isometry of grid cells: Learning distance-preserving position embedding. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=Xo0Q1N7CGk. Oral presentation.

Tie Xu, Shengdun Wu, Junwen Luo, Yao Yang, Feng Lin, Junjie Jiang, Guozhang Chen, and Dongping Yang. Representation transfer via invariant input-driven continuous attractors for fast domain adaptation. *Communications Biology*, 9:747, 2026. doi: 10.1038/s42003-026-09938-8.

Wenhao Zhang, Ying Nian Wu, and Si Wu. Translation-equivariant representation in recurrent networks with a continuous manifold of attractors. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/65384a01325fecbd364c835db872443c-Abstract-Conference.html.

Junfeng Zuo, Xiao Liu, Ying Nian Wu, Si Wu, and Wenhao Zhang. A recurrent neural circuit mechanism of temporal-scaling equivariant representation. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/c5e44243e16c9d61d3897ba1095f5f6c-Abstract-Conference.html.

Table 1: Three questions about the same continuous memory. Each analysis has explicit assumptions; complete proofs and the technical corollaries are in the appendix.

Question	What to measure	What it tells us
Which differences survive?	Independent state directions generated by the symmetry	Protected dimension, assuming exact input-wise symmetry and nondegenerate representation
How do imperfections act?	Direct damage, outward and return couplings, and transverse contraction	Sufficient local bounds and offline bounds along a supplied finite trajectory
How does noise affect precision?	Representation sensitivity, decoder sensitivity, and state-noise covariance	Local decoded covariance and geometry choices under a declared resource budget

A Exact group-tangent transport

A.1 Proof of the controlled-map theorem

Proof of Theorem 1. For any $\xi \in \mathfrak{g}$, per-input equivariance gives

$$F_u(\exp(s\xi) \cdot x) = \exp(s\xi) \cdot F_u(x).$$

Differentiating at $s = 0$ yields

$$DF_u(x)A_x\xi = A_{F_u(x)}\xi. \tag{20}$$

Iterating Equation (20) along $x_n, \dots, x_m$ proves Equation (3). Equivariance implies stabilizer monotonicity: if $g \cdot x_t = x_t$, then $g \cdot x_{t+1} = F_{u_t}(g \cdot x_t) = F_{u_t}(x_t) = x_{t+1}$. Hence $\ker A_{x_t} \subseteq \ker A_{x_{t+1}}$. Since every A_x has rank q on K, the kernels have equal dimension and therefore coincide along the trajectory. The conclusion used below is equality of the isotropy Lie algebras $\ker A_{x_t}$ along this trajectory. Under a proper action, full stabilizers within one orbit-type stratum are conjugate and need not be literally equal as embedded subgroups.

Write

$$P_{m:n} := DF_{u_{m-1}}(x_{m-1}) \cdots DF_{u_n}(x_n).$$

For $v \in E^G_{x_n}$, choose the unique $\xi \in (\ker A_{x_n})^\perp$ such that $v = A_{x_n}\xi$. Kernel equality makes $\xi \in (\ker A_{x_m})^\perp$, and exact transport gives $P_{m:n}v = A_{x_m}\xi$. The nonzero-singular-value bounds therefore imply

$$\frac{c}{C}\|v\| \le \|P_{m:n}v\| \le \frac{C}{c}\|v\|$$

uniformly in $m - n$. Thus the operator norm and conorm of the restricted cocycle are uniformly bounded, and every restricted singular-value growth rate is squeezed between $(m-n)^{-1}\log(c/C)$ and $(m-n)^{-1}\log(C/c)$ and tends to zero. For a prescribed nonstationary input sequence, these uniform bounds still hold pathwise, but no stationary Oseledets spectrum is asserted unless the required limits exist. □

A.2 Stationary full-spectrum inclusion

Corollary 2 (Stationary controlled cocycle)**.** *Let $(\Omega, \mathcal{F}, \nu)$ be a standard Borel probability space, let ϑ be invertible, ergodic, and measure preserving, and set $Y = \Omega \times K$. Suppose $\Theta(\omega, x) = (\vartheta\omega, F_\omega(x))$ is measurable and preserves an ergodic probability measure μ on Y whose Ω-marginal is ν. For the full-spectrum assertion, either assume that Θ is invertible modulo μ, or pass to the natural extension of (Y, μ, Θ) and lift the cocycle and group-tangent bundle through its zeroth-coordinate projection. Assume the derivative cocycle $L(\omega, x) = DF_\omega(x)$ and E^G_x are measurable and*

$$\int_{\Omega \times K} \log^+\|L(\omega, x)\| \, \mathrm{d}\mu(\omega, x) < \infty. \tag{4}$$

Under the hypotheses of Theorem 1, the restricted cocycle on the invariant rank-q bundle E^G has q zero Lyapunov exponents for μ-almost every base point. On the invertible skew product or its natural extension, a measurable adapted orthonormal frame makes the full derivative cocycle block upper

triangular with this restricted cocycle as a diagonal block. The finite-dimensional semi-invertible multiplicative ergodic theorem therefore implies that the full forward spectrum contains at least q *zero exponents for μ-almost every base point, counted with multiplicity (Oseledets, 1968; Froyland et al., 2013). Additional flow-neutral or transverse zero exponents are not excluded.*

Proof of Corollary 2. Write $Y = \Omega \times K$. If Θ is invertible modulo μ, work on (Y, μ, Θ) itself. Otherwise let $(\widehat{Y}, \widehat{\mu}, \widehat{\Theta})$ be its natural extension, with zeroth-coordinate projection π_0 satisfying $\pi_0\widehat{\Theta} = \Theta\pi_0$ and $(\pi_0)_*\widehat{\mu} = \mu$. Define $\widehat{L}(\widehat{y}) = L(\pi_0\widehat{y})$ on the pulled-back tangent fibers and $\widehat{E}^G_{\widehat{y}} = E^G_{\pi_0\widehat{y}}$. Then

$$\widehat{L}^{(n)}(\widehat{y}) = L^{(n)}(\pi_0\widehat{y}),$$

so lifting changes neither forward products nor their Lyapunov exponents. The chosen base transformation is now invertible; only the linear fiber maps may remain noninvertible, as allowed by the semi-invertible theorem. All statements below hold almost everywhere with respect to the relevant invariant measure.

The rank-q bundle $\widehat{E}^G$ is measurable by assumption. Exact transport and constant rank give $\widehat{L}(\widehat{y})\widehat{E}^G_{\widehat{y}} = \widehat{E}^G_{\widehat{\Theta}\widehat{y}}$, so the restriction is an invariant linear cocycle. The pathwise bounds in Theorem 1 imply that every nonzero restricted vector has exponent zero and that all q restricted singular-value exponents are zero.

Choose a measurable complement of $\widehat{E}^G$. In the resulting adapted frames the full cocycle has the block form

$$\begin{pmatrix} L_G & * \\ 0 & L_Q \end{pmatrix},$$

where L_G is the restriction to $\widehat{E}^G$ and L_Q is the induced quotient cocycle. The integrability in Equation (4) also controls these finite-dimensional diagonal blocks. The invariant-subbundle property of the semi-invertible multiplicative ergodic theorem, equivalently the spectrum union property for this block-upper-triangular cocycle, places the q zero exponents of L_G in the full spectrum with multiplicity (Froyland et al., 2013). Thus the full spectrum contains at least q zeros; the theorem does not exclude additional flow-neutral or transverse zeros. Because the lifted and original forward products coincide under π_0, this is also the full forward spectrum over μ. No statement about a stationary full spectrum is made for an arbitrary prescribed control sequence. □

Lemma 8 (Uniform nondegeneracy on a compact constant-rank stratum)**.** *Let K be compact and $x \mapsto A_x$ continuous, where every $A_x : \mathfrak{g} \to T_xM$ has rank q. Then there exist $0 < c \le C < \infty$ such that all nonzero singular values of A_x lie in $[c, C]$ for $x \in K$.*

Proof. The largest singular value and the smallest nonzero singular value are continuous on a constant-rank family in fixed local orthonormal frames. The former attains a finite maximum on K, while rank q makes the latter positive pointwise and compactness makes its minimum positive. Equivalent local trivializations give the same conclusion because transition maps are uniformly bounded on compact overlaps. □

The explicit conditioning assumption in the main text is therefore redundant for a smooth action on a compact constant-rank stratum, but it is useful computationally: C/c and the reducing-frame condition number quantify how a small slope in adapted coordinates can appear in the ambient Euclidean metric.

A.3 Autonomous-flow corollary and the ordinary flow-zero mode

Corollary 9 (Autonomous equivariant flow)**.** *Let $\dot{x} = f(x)$ be a complete C^1 flow ϕ_t satisfying $f(g \cdot x) = D(g)_x f(x)$. Under the compact constant-rank hypotheses of Theorem 1,*

$$D\phi_t(x)A_x = A_{\phi_t(x)}, \tag{21}$$

and every nonzero group-tangent vector has zero direct logarithmic growth rate. Applying the stationary hypotheses of Corollary 2 to a fixed positive-time map, the full Oseledets spectrum contains at least q zero exponents, counted with multiplicity.

Proof. Equivariance of the vector field and uniqueness of solutions imply flow equivariance. For every $\xi \in \mathfrak{g}$,

$$\phi_t(\exp(s\xi) \cdot x) = \exp(s\xi) \cdot \phi_t(x).$$

Differentiating this identity at $s = 0$ gives

$$D\phi_t(x)\xi_M(x) = \xi_M(\phi_t(x)),$$

which is Equation (21). Restrict ξ to a fixed linear complement of $\ker A_x$; its dimension is $q = \dim \mathfrak{g} - \dim \ker A_x = \dim(G/H)$, and the restricted infinitesimal action is injective. The uniform nonzero-singular-value bounds then squeeze every direct group-tangent growth rate to zero exactly as in Theorem 1. The full-spectrum conclusion is the separate MET corollary, not a consequence of this differentiation alone. This is the standard infinitesimal equivariance identity used in equivariant dynamics (Golubitsky et al., 1988; Krupa, 1990). □

For an invertible flow, the full stabilizer is preserved exactly along an orbit: $G_x \subseteq G_{\phi_t(x)}$ follows from equivariance, and applying the same argument to ϕ_{-t} gives equality. The ordinary autonomous-flow zero mode is generated by $f(x)$. It may coincide with one group tangent at a relative equilibrium, but it vanishes on a fixed-point continuous attractor. Therefore, a single near-zero exponent is not sufficient evidence for a protected memory mode. Product groups with $q > 1$, direct analytical tangents, and principal-angle alignment resolve this ambiguity.

A.4 Why a finite symmetry group is different

A finite group can enforce parameter sharing, block structure, and spectral degeneracies, but its Lie algebra is zero. It therefore provides no nonzero fundamental vector fields ξ_M and does not imply a continuously parameterized neutral tangent through Equation (20). A finite circulant grid has exact integer-roll C_N equivariance but generally fails equivariance under a generic continuous phase shift. Such grids remain useful approximations and null controls, but they are not theorem models for continuous memory.

B Additional exact-geometry diagnostics

The overview in Table 1 summarizes the role of these additional diagnostics. The numerical thresholded count is always secondary to direct tangent transport and analytical subspace alignment.

Table 2: Representative exact-symmetry diagnostics. Values are maxima over the corresponding evaluations fixed before analysis unless otherwise stated. The learned S^1 step residual is inherited from the original float32 task implementation; the native-coordinate perturbation analysis reloads the same checkpoints in float64 and therefore reports roundoff-scale exact-symmetry residuals. The values reflect different floating-point evaluation regimes.

Diagnostic	Value
Coupled-irrep vector-field equivariance error	6.97×10^{-16}
Coupled-irrep direct group-tangent exponent magnitude	5.30×10^{-13}
Coupled-irrep tangent-covariance angle	1.48×10^{-6} degrees
Broken coupled-irrep equivariance error	1.29×10^{-2}
Learned S^1 maximum exact step-equivariance error	3.19×10^{-8}
Learned S^1 mean zero-input direct tangent exponent	-2.28×10^{-6}
Learned S^1 mean zero-input principal angle	2.97×10^{-2} degrees

C Blockwise invariant-graph criterion and rate bounds

C.1 Reducing an upper-triangular exact map

Let $X \oplus N$ be a finite-dimensional tangent/normal splitting and suppose an exact local time map has derivative

$$T_0 = \begin{pmatrix} I & H \\ 0 & B \end{pmatrix}. \tag{22}$$

The lower-left zero follows from invariance of the group tangent; the upper-right block depends on the chosen complement.

Lemma 10 (Complement-shear reduction)**.** *If $I - B$ is invertible, set*

$$C = -H(I - B)^{-1}, \qquad S = \begin{pmatrix} I & C \\ 0 & I \end{pmatrix}. \tag{23}$$

Then $S^{-1}T_0S = \mathrm{diag}(I, B)$. If $\|B\| < 1$, $I - B$ is invertible by the Neumann series.

Proof. Direct multiplication gives the transformed upper-right block $H + C(I - B) = 0$. The diagonal blocks and lower-left zero are unchanged. □

The experiment records the map duration, the normal gap, the shear reduction residual, and $\kappa_S = \|S\|\|S^{-1}\|$. This prevents an apparently small adapted-coordinate angle from being presented without its ambient conditioning cost.

Lemma 11 (Directed gap equals the sine of the largest principal angle)**.** *Let U, V be equal q-dimensional subspaces of a finite-dimensional Euclidean space, with $q \geq 1$. If P_V is the orthogonal projector onto V and $0 \leq \theta_1 \leq \cdots \leq \theta_q \leq \pi/2$ are their principal angles, then*

$$\delta(U, V) := \sup_{\substack{u \in U \\ \|u\|=1}} \mathrm{dist}(u, V) = \|P_{V^\perp}|_U\| = \sin\theta_q. \tag{24}$$

Thus $\arcsin\delta(U, V) = \theta_q$, the conventional largest principal angle.

Proof. The nearest point in V to u is $P_V u$, so $\mathrm{dist}(u, V) = \|(I - P_V)u\|$, proving the first equality. Let Q_U, Q_V have orthonormal columns spanning U, V. By the singular-value definition of principal angles, the singular values of $Q_V^\top Q_U$ are $\cos\theta_i$ (Knyazev & Argentati, 2002). Moreover,

$$Q_U^\top(I - Q_V Q_V^\top)Q_U = I - (Q_V^\top Q_U)^\top(Q_V^\top Q_U).$$

Its eigenvalues are therefore $\sin^2\theta_i$. Hence the operator norm of $(I - P_V)Q_U$, equivalently of $P_{V^\perp}|_U$, is $\max_i \sin\theta_i = \sin\theta_q$. □

C.2 Proof of Proposition 3

For a graph $\{(x, Lx) : x \in X\}$, the image under Equation (7) is a graph whenever the center factor is invertible, with graph transform

$$\Phi(L) = \big(E_{NG} + (B + E_{NN})L\big)\big(I + E_{GG} + E_{GN}L\big)^{-1}. \tag{25}$$

If $\|L\| \leq r$ and $a + br < 1$, the inverse exists and

$$\|\Phi(L)\| \leq \frac{c + (\rho + e)r}{1 - a - br}. \tag{26}$$

The right-hand side is at most r exactly when

$$br^2 - (d - a - e)r + c \leq 0. \tag{27}$$

Under $D = d - a - e > 0$ and $D^2 \geq 4bc$, the smaller nonnegative root is Equation (8); the displayed form is numerically stable when bc is small. For $b = 0$, Equation (27) reduces to $c - Dr \leq 0$ and $r_* = c/D$. The graph ball is compact and convex in the finite-dimensional operator space, and Φ is continuous on it, so Brouwer's theorem supplies at least one fixed graph. This existence argument does not claim uniqueness.

On a fixed graph, write $J_L x = (x, Lx)$ and $M_L = I + E_{GG} + E_{GN}L$. The restricted map is conjugate to its center-coordinate representation, $T|_{\Gamma_L} = J_L M_L J_L^{-1}$. The distance from M_L to identity is at most $s = a + br_*$, so its norm is at most $1 + s$ and its conorm at least $1 - s$. In center coordinates, k iterates of the same local map obey

$$(1 - s)^k\|x\| \leq \|M_L^k x\| \leq (1 + s)^k\|x\|.$$

Passing to the ambient graph norm introduces only the fixed norm-equivalence factor

$$\kappa(J_L)^{-1}(1-s)^k\|v\| \le \|T^k v\| \le \kappa(J_L)(1+s)^k\|v\|, \qquad v \in \Gamma_L.$$

Because $\kappa(J_L)$ is independent of k, its logarithm vanishes after division by physical time $k\tau$ and passage to the asymptotic rate. This proves Equation (9). The argument concerns repeated iterations of the same selected local linear map; it does not establish a time-varying graph-sequence theorem.

To obtain the ambient-angle bound, a reducing-coordinate graph vector (x, Lx) lies at distance at most $\|S\|\, r_*\|x\|$ from $S(X, 0)$, while

$$\|S(x, Lx)\| \ge \|S^{-1}\|^{-1}\|(x, Lx)\| \ge \|S^{-1}\|^{-1}\|x\|.$$

Taking the supremum over unit graph vectors bounds the directed gap from the ambient graph to the exact center plane by $\min\{1, \kappa_S r_*\}$. The two spaces have the same dimension, so Lemma 11 gives $\sin\alpha \le \min\{1, \kappa_S r_*\}$ for their largest principal angle.

C.3 Perturbative orders and one-way leakage

With $a = e = 0$, write $D = d$ and expand the smaller root:

$$r_* = \frac{2c}{d + d\sqrt{1 - 4bc/d^2}} = \frac{c}{d} + \frac{bc^2}{d^3} + O((bc)^2 c/d^5), \tag{28}$$

$$s = br_* = \frac{bc}{d} + O(b^2c^2/d^3). \tag{29}$$

Thus, with $d > 0$ fixed as $\epsilon \to 0$, $b, c = O(\epsilon)$ imply $s = O(\epsilon^2/d)$. If $c = 0$, then $r_* = 0$ and upper leakage b alone leaves the exact center graph invariant. If $b = 0$, the graph tilts by at most c/D, but the center map remains $I + E_{GG}$; lower leakage alone does not change the center exponent in the constant local block model. Normal-only perturbations similarly leave the center exponent at numerical zero, which is why bound/measured ratios are not reported for that family.

C.4 Local affine residual realization

Lemma 12 (Realizing a prescribed local Jacobian defect)**.** *Let $F_0 : \mathbb{R}^n \to \mathbb{R}^n$ be C^1, let x_* be a selected orbit point, and let $\Delta \in \mathbb{R}^{n\times n}$. There exists a compactly supported C^1 residual branch R such that*

$$R(x_*) = 0, \qquad DR(x_*) = \Delta.$$

Hence $F_\epsilon = F_0 + \epsilon R$ has the same selected state value and local Jacobian $DF_0(x_*) + \epsilon\Delta$.

Proof. Choose a C^1 bump function χ equal to one in a neighborhood of x_* and define $R(x) = \chi(x)\Delta(x - x_*)$. Then $R(x_*) = 0$ and $DR(x_*) = \Delta$. □

This lemma justifies calling the scalar selected-map study a causal local recurrent residual intervention. It does not imply that an arbitrary weight perturbation has the same block routing, nor that one local branch preserves any global symmetry or orbit away from the selected point.

D Finite-horizon bounds along controlled recurrent trajectories

Scope and coherent coordinates. Fix $n < m$. Let $X \cong \mathbb{R}^q$ and $N \cong \mathbb{R}^p$ be fixed Euclidean model spaces over the prescribed finite horizon, and let $S_t : X \oplus N \to T_{x_t}M$ be invertible for every endpoint. The same endpoint frame is used as the target frame at step $t-1$ and the source frame at step t. The coordinate cocycle is

$$T_t = S_{t+1}^{-1} DF_{u_t}(x_t) S_t. \tag{30}$$

Let $A_t : \mathfrak{g} \to T_{x_t}M$ be the infinitesimal-action map. Exact per-input equivariance gives

$$DF_{u_t}(x_t) A_t = A_{t+1}. \tag{31}$$

Assume $\operatorname{rank} A_t = q$ along the trajectory. Then $\ker A_t \subseteq \ker A_{t+1}$, and equality of dimensions makes these kernels one common subspace $\mathfrak{h}_* \subset \mathfrak{g}$. Fix once and for all a complement V with $\mathfrak{g} = \mathfrak{h}_* \oplus V$ and a basis of V (equivalently, a fixed basis of the vector-space quotient $\mathfrak{g}/\mathfrak{h}_*$; no quotient Lie algebra is asserted). Use the q full-rank columns of $A_t|_V$ at every endpoint. Passing all raw Lie-algebra columns is valid only for a locally free action; with a nontrivial stabilizer those columns are rank deficient. Endpoint-wise QR normalization generally inserts nonidentity transition matrices, which cannot be silently replaced by an identity center block. Appending orthogonal normal complements first gives

$$\begin{pmatrix} I & C_t \\ 0 & B_t \end{pmatrix}. \tag{32}$$

With $R_t = \left(\begin{smallmatrix} I & K_t \\ 0 & I \end{smallmatrix}\right)$, direct multiplication gives

$$R_{t+1}^{-1} \begin{pmatrix} I & C_t \\ 0 & B_t \end{pmatrix} R_t = \begin{pmatrix} I & K_t + C_t - K_{t+1} B_t \\ 0 & B_t \end{pmatrix}. \tag{33}$$

Thus the terminal anchor $K_m = 0$ and backward recurrence $K_t = K_{t+1} B_t - C_t$ remove C_t on the declared horizon. Explicitly,

$$K_t = -\sum_{j=t}^{m-1} C_j B_{j-1} \cdots B_t, \qquad \|K_t\| \le \sum_{j=t}^{m-1} \|C_j\| \prod_{\ell=t}^{j-1} \rho_\ell, \tag{34}$$

where the product is the identity when $j = t$. Hence, if $\rho_t \le \bar{\rho} < 1$ and $\|C_t\| \le \bar{C}$, then $\|K_t\| \le \bar{C}(1 - \bar{\rho}^{m-t})/(1 - \bar{\rho})$; without a uniform gap the shear may grow with the horizon. The shear leaves the action columns unchanged but depends on the future reference sequence; the resulting reduction is therefore offline/trajectory-conditioned unless the future controls and reference maps are known. Its horizon and gap dependence is retained in S_t and $\kappa(S_t)$ rather than discarded.

In these reducing frames, write explicitly

$$T_t = \begin{pmatrix} I + E_{GG,t} & E_{GN,t} \\ E_{NG,t} & B_t + E_{NN,t} \end{pmatrix}, \quad \begin{matrix} \|E_{GG,t}\| \le a_t, & \|E_{GN,t}\| \le b_t, \\ \|E_{NG,t}\| \le c_t, & \|E_{NN,t}\| \le e_t, \end{matrix} \tag{35}$$

with $\|B_t\| \le \rho_t < 1$.

Proof of Proposition 4. For $J_L v = (v, Lv)$, block multiplication gives

$$M_t v = (I + E_{GG,t} + E_{GN,t} L_t) v, \tag{36}$$

$$L_{t+1} = \{E_{NG,t} + (B_t + E_{NN,t}) L_t\} M_t^{-1}. \tag{37}$$

If $\|L_t\| \le r_t$, then $\|M_t - I\| \le a_t + b_t r_t = s_t < 1$, so M_t is invertible and $\|M_t^{-1}\| \le (1 - s_t)^{-1}$. Taking norms in Equation (37) gives exactly Equation (11); induction proves graph transport and $\|L_t\| \le r_t$. The same multiplication gives $T_t J_{L_t} = J_{L_{t+1}} M_t$.

Each $M_t : X \to X$ has conorm at least $1 - s_t$ and norm at most $1 + s_t$. For the chronological center product $P_{m:n} = M_{m-1} \cdots M_n$, supermultiplicativity of conorm and submultiplicativity of norm give

$$\prod_{t=n}^{m-1} (1 - s_t) \le \sigma_{\min}(P_{m:n}) \le \|P_{m:n}\| \le \prod_{t=n}^{m-1} (1 + s_t). \tag{38}$$

Multiplying the graph identities telescopes to

$$DF_{m:n} G_n = G_m P_{m:n}, \qquad G_t = S_t J_{L_t}, \tag{39}$$

where $DF_{m:n} = DF_{u_{m-1}}(x_{m-1}) \cdots DF_{u_n}(x_n)$. Thus the lower ambient multiplier is bounded by $[\sigma_{\min}(G_m)/\|G_n\|] \prod_t (1 - s_t)$ and the upper multiplier by $[\|G_m\|/\sigma_{\min}(G_n)] \prod_t (1 + s_t)$. Directly computable endpoint bounds are

$$\sigma_{\min}(G_t) \ge \|S_t^{-1}\|^{-1}, \qquad \|G_t\| \le \|S_t\| \sqrt{1 + r_t^2}. \tag{40}$$

Consequently, for every nonzero $v \in X$, positive step durations τ_t, and $\Delta\tau = \sum_{t=n}^{m-1} \tau_t$,

$$\frac{1}{\Delta\tau} \log \left[\frac{\sigma_{\min}(G_m)}{\|G_n\|} \prod_{t=n}^{m-1} (1 - s_t) \right] \le \frac{1}{\Delta\tau} \log \frac{\|DF_{m:n} G_n v\|}{\|G_n v\|}, \tag{41}$$

$$\frac{1}{\Delta\tau} \log \frac{\|DF_{m:n} G_n v\|}{\|G_n v\|} \le \frac{1}{\Delta\tau} \log \left[\frac{\|G_m\|}{\sigma_{\min}(G_n)} \prod_{t=n}^{m-1} (1 + s_t) \right]. \tag{42}$$

Finally, the distance from $S_t(v, L_t v)$ to the reference center plane is at most $\|S_t\| r_t \|v\|$, whereas its norm is at least $\sigma_{\min}(S_t)\|v\|$. Taking the worst normalized distance bounds the directed gap by $\min\{1, \kappa(S_t) r_t\}$. Since the transported graph and reference center plane both have dimension q, Lemma 11 gives $\sin \alpha_t \le \min\{1, \kappa(S_t) r_t\}$ for the conventional largest principal angle. Only endpoint frame factors remain because the same coherent sequence appears on both sides of every adjacent map. This proves the finite-horizon statement for the one specified initial graph. It does not select that graph, prove existence or uniqueness of a bi-infinite invariant bundle, or identify an asymptotic Lyapunov spectrum. Such a continuation would require stronger exponential-dichotomy or normal-hyperbolicity hypotheses and belongs to the classical persistence framework (Fenichel, 1971; Hirsch et al., 1977).

Formalization. The finite-sequence graph transport, chronological product bounds, endpoint factors, terminal shear, and leakage estimate are machine-checked in Lean 4. Lean proves directed-gap bounds; Lemma 11 identifies the conventional largest principal angle in the manuscript. The endpoint-indexed construction does not require smooth dependence on the base point. Section N.1 gives the complete scope and separates these results from the general Lie-orbit and multiplicative-ergodic interfaces.

Constant-map and perturbative limits. For a repeated map, initialize the scalar recurrence at r_* from Equation (8). Proposition 3 supplies an invariant graph in the corresponding closed ball. Because

$$c + (\rho + e) r_* = r_*(1 - a - b r_*),$$

the scalar radius remains r_*, $s_t = a + b r_*$, and the products reduce to $(1 - s)^k$ and $(1 + s)^k$. Starting from zero instead describes a transient forward graph.

The leakage order can be stated quantitatively. Suppose $r_n = 0$, $a_t = e_t = 0$, $\rho_t \le 1 - d$ with $0 < d \le 1$, and $\epsilon \ge 0$. Uniformly in t, let $b_t \le B\epsilon$ and $c_t \le C\epsilon$ for constants $B, C \ge 0$. If $4BC\epsilon^2 \le d^2$, then induction with $R = 2C\epsilon/d$ gives

$$r_t \le R, \qquad 1 - b_t r_t \ge 1 - d/2 > 0, \qquad s_t = b_t r_t \le \frac{2BC\epsilon^2}{d} \tag{43}$$

for every step on the declared sequence. Indeed, $C\epsilon + (1 - d)R \le R(1 - B\epsilon R)$ is equivalent to $C\epsilon \le dR - B\epsilon R^2$, which follows from the stated smallness condition. Thus bidirectional leakage is second order in the center-coordinate rate bound under these uniform hypotheses, while direct center damage $a_t = O(\epsilon)$ enters s_t at first order. Ambient rates also retain the endpoint chart terms in Equation (42); no asymptotic conclusion follows without controlling those terms and the frame sequence as the horizon grows.

Controlled-trajectory evaluation. We tested the finite-horizon bounds on eight held-out base models, each with three input trajectories. The 32-step macro-map protocol was fixed before evaluation, and no nonlinear endpoint metric was used for selection. No violations of the coordinate, ambient, or angle bounds were observed in the 216 structured-residual cases satisfying the assumptions. We also retained 24 controls constructed to violate the assumptions; all were identified as failing the sufficient conditions. We first averaged the fitted scaling exponents across the three trajectories within each model, then computed the mean and SEM across the eight models. The estimates were 0.916 ± 0.095 for direct damage and 2.040 ± 0.092 for leakage (mean $\pm$ SEM), consistent with the predicted first- and second-order scaling over the tested finite horizon. These results assess predictions made before finite-time dynamics for the specified structured, time-varying residuals; they do not establish generality across errors arising from optimization, quantization, or hardware.

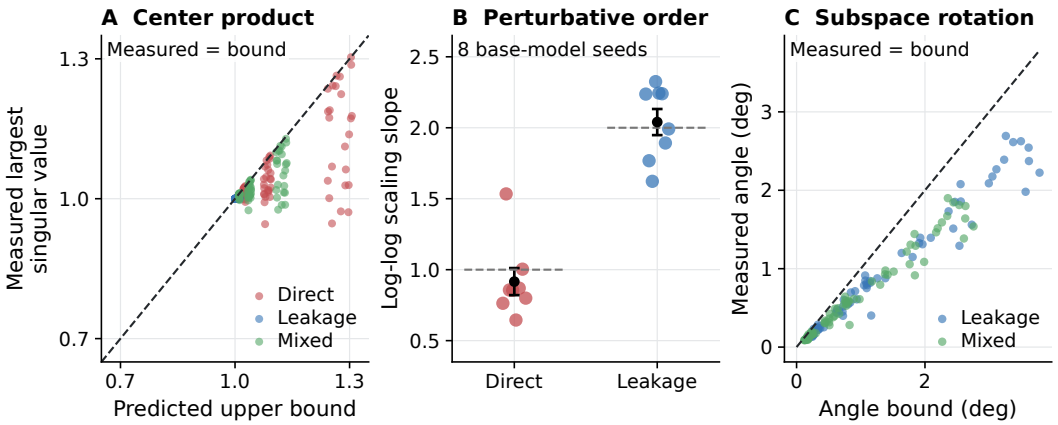


Figure 6: **Finite-horizon results along controlled trajectories.** (A) Coordinate-product upper bounds versus measured largest singular values. (B) One point per base-model seed after averaging its nested trajectories; black marks are mean ± SEM over eight seeds, and dashed references are orders one and two. (C) Bounds versus measured subspace angles. Dashed diagonals in A/C mark equality; points below them respect the upper bound. The 24 assumption-violation controls remain in the released tables. This offline, finite-horizon test of structured residuals gives no asymptotic, bi-infinite, online, or universal-perturbation guarantee.

E Selected-map perturbation study

E.1 Models, maps, and interventions

The experiment fixed six coupled-irrep recurrent instances and all six learned exact S^1 checkpoints. For each model, an exact local time map at a fixed orbit point is reduced using Lemma 10. Perturbations are inserted in the reducing blocks and transformed back to ambient coordinates. The protected bundle is scalar ($q = 1$) in every instance of this study; the proposition itself permits general finite q. The intervention families are:

- direct tangent: $E_{GG} \neq 0$ only;
- bidirectional leakage: $E_{GN}, E_{NG} \neq 0$ with $E_{GG} = E_{NN} = 0$;
- normal only: $E_{NN} \neq 0$ only;
- mixed: all blocks active under the routing rule fixed before evaluation.

Seven scales $\epsilon \in \{10^{-5}, 3\times10^{-5}, 10^{-4}, 3\times10^{-4}, 10^{-3}, 3\times10^{-3}, 10^{-2}\}$ are evaluated. Directional replication creates nested rows within a model and scale; statistical summaries therefore first aggregate to the model level.

Table 3: Local map and frame ranges for the scalar perturbation study.

Family	Models	Map duration τ	Normal gap d	κ_S
Coupled-irrep RNN	6	0.16 (five), 0.32 (one)	0.1206–0.2086	4.7435
Learned equivariant cell	6	0.8	0.1682–0.1741	1.0000

The direct perturbations use the same analytic construction in every model, which fixes the leading linear term. This explains the nearly identical fitted slopes and small SEM. The results support the numerical implementation of the prescribed direct perturbation, but do not establish a universal first-order coefficient for other implementation errors.

E.2 Conditions and bound coverage

Of 1,608 total rows, 12 are zero-defect references and 1,596 are nonzero. The sufficient conditions hold in 1,572 nonzero rows. The only failing nonzero rows are the 24 learned mixed-direction cases

at $\epsilon = 10^{-2}$. A total of 1,476 rows have informative exponent and ambient-angle bounds. The additional uninformative rows are 48 coupled leakage and 48 coupled mixed cases at $\epsilon = 10^{-2}$. Among all rows satisfying the conditions, the observed counts of exponent, invariant-graph-radius, and angle violations are each zero.

These are coverage counts over a structured test set, not independent Bernoulli trials. Scaling uncertainties use the 12 model instances rather than treating their perturbation rows as independent replicates.

Table 4: Perturbative slopes, fit separately within each model and summarized across six models per family.

Model family	Defect family	Mean slope	SEM	Mean R^2
Coupled-irrep	direct tangent	1.000575	2.5×10^{-12}	0.9999998
Learned cell	direct tangent	1.000575	5.4×10^{-13}	0.9999998
Coupled-irrep	bidirectional leakage	2.003847	0.000131	0.9999951
Learned cell	bidirectional leakage	2.023140	0.000729	0.9998185

Table 5: Sharpness of nontrivial physical-time exponent bounds. Ratios are computed only when the measured exponent is above the numerical floor.

Model family	Defect family	Median bound/measured	10th pct.	90th pct.
Coupled-irrep	direct tangent	1.000	1.000	1.000
Coupled-irrep	leakage	1.468	1.329	1.698
Coupled-irrep	mixed	1.005	1.000	1.230
Learned cell	direct tangent	1.000	1.000	1.000
Learned cell	leakage	1.036	1.022	1.116
Learned cell	mixed	1.000	1.000	1.013

F NATIVE-PARAMETER AND AMBIENT-RESIDUAL PERTURBATIONS

In the scalar-memory study, we prescribe perturbations in the reducing coordinates of the selected map to separate direct tangent effects from transverse feedback. Here, we instead introduce perturbations in the recurrent model's original coordinates and examine how they contribute to the tangent and transverse blocks of the selected-map perturbation. We use all six learned checkpoints of the exactly S^1-equivariant architecture, with the same zero-input orbit point and selected time maps.

Exact-parameterization control. We perturb every trainable tensor with independent Gaussian noise at five relative perturbation scales ϵ between 10^{-4} and 10^{-2}, using three draws per scale and checkpoint. For a tensor W, the noise standard deviation is $\epsilon \max\{\mathrm{RMS}(W), 10^{-3}\}$, where $\mathrm{RMS}(W)$ is the root-mean-square value of its entries. Because the architecture constructs its coefficients only from invariant features, all 90 perturbed models retain exact equivariance in real arithmetic. The maximum sampled step-equivariance error is 1.121×10^{-16}, the maximum analytical tangent-transport residual is 1.241×10^{-16}, and the largest measured protected-exponent magnitude is 8.33×10^{-16}. Thus ordinary parameter error inside a hard-equivariant parameterization is not itself a symmetry defect, although it can change transverse dynamics.

Ambient residual parameters. For each checkpoint we draw 20 unit-operator-norm matrices from each of three families: dense Gaussian, 10%-dense sparse Gaussian, and rank-one outer products. At the selected state x_*, the local residual branch is $R_P(x) = \chi(x)P(x - x_*)$ as in Lemma 12. Seven scales give 2,520 nonzero cases. The residual is specified in ambient state coordinates; only afterward is its induced selected-time-map defect projected into $(E_{GG}, E_{GN}, E_{NG}, E_{NN})$. The bound is therefore evaluated without using the measured perturbed invariant mode.

Of 2,520 cases, 2,515 satisfy the conditions and 2,514 have informative exponent and angle bounds. No case satisfying the conditions violates the exponent, graph, or angle bound. The five failing

and one additional angle-uninformative cases all occur at $\epsilon = 10^{-2}$ and are retained. The block-resolved bound is more descriptively associated with the measured displacement than the scalar local generator-commutator norm (log–log Pearson $r = 0.834$ versus 0.533), but these correlations are not inferential because scales and directions are nested within checkpoint.

Table 6: Direct and transverse-feedback effects of native-coordinate residuals. The direct-dominant fraction compares the direct tangent perturbation magnitude $a = \|E_{GG}\|_2$ with the leading-order feedback proxy bc/d. Here, $c = \|E_{NG}\|_2$ measures coupling from memory into transverse directions, $b = \|E_{GN}\|_2$ measures coupling back into memory, and $d = 1 - \rho$ is the reference transverse contraction margin, with $\|B\|_2 \le \rho < 1$. These quantities are computed from the local linearization, before the multi-step dynamics are evaluated, and compare local perturbation magnitudes rather than contributions to measured task error. Fitted scaling exponents are first summarized over perturbation directions within each checkpoint and then reported as mean ± SEM across six checkpoints. For sparse residuals, structural zeros produce a mixture of perturbative orders rather than a single universal scaling exponent.

Residual family	Cases	Valid	Direct-dominant fraction	Checkpoint slope, mean ± SEM
Dense	840	99.64%	96.90%	1.0028 ± 0.0031
Sparse	840	99.76%	40.83%	1.9117 ± 0.2715
Rank one	840	100.00%	100.00%	1.0001 ± 0.0016

G Local predictions and multi-step dynamics in a two-coordinate memory

We ask whether the local perturbation analysis remains informative when applied to a learned memory with two continuous coordinates and when the perturbation is introduced directly into the recurrent update. Unlike the preceding controlled study, which specifies perturbations in reducing coordinates, this experiment perturbs the recurrent dynamics first and only then decomposes the resulting selected-map perturbation into memory and transverse blocks.

We use the six condition-one T^2 checkpoints from the prescribed-geometry study and evaluate the memory state with phases $(0.37, -0.41)$. For each checkpoint, the unperturbed one-step Jacobian has two identity group-tangent directions. We compose eight recurrent steps to form the selected map used in the local analysis. Across checkpoints, the transverse contraction gaps range from 0.1646 to 0.1919, the reducing-frame condition number is one, and the fixed-state, identity, invariance, and reduction residuals are at numerical precision.

The broken update is

$$F_{\epsilon,P}(x,u) = F_0(x,u) + \epsilon P(x - x_*), \tag{44}$$

where $\|P\|_2 = 1$ and P is a channelwise gain-mismatch matrix, a rank-two negative-semidefinite adapter, or a dense negative-semidefinite residual adapter. The anchor keeps x_* fixed, so its selected-map Jacobian is computed exactly from $(DF_0(x_*) + \epsilon P)^8$. These ambient-coordinate adapters break T^2 equivariance and generically populate several reducing blocks. They are representative tying/residual errors, not a measured distribution from a specific optimizer, quantizer, or device.

For a locally pinning continued tangent plane, let ζ_1, ζ_2 be the selected-map eigenvalues for map duration τ, and define the two positive decay rates by $\gamma_j = -\tau^{-1} \log |\zeta_j|$. Order them as $\gamma_{\text{fast}} \ge \gamma_{\text{slow}} > 0$. We summarize the two local decay modes by the scalar spectral half-life

$$\tau_{1/2}^{\text{geom}} = \sqrt{\frac{\log 2}{\gamma_{\text{fast}}} \frac{\log 2}{\gamma_{\text{slow}}}}, \tag{45}$$

which is the geometric mean of the fast- and slow-mode half-lives. We compare this local prediction with the median half-life measured from the subsequent multi-step dynamics. The individual fast- and slow-mode half-lives are also reported in the supplementary table.

For that comparison, 16 probes begin at Euclidean radius 0.08 in wrapped two-phase coordinates around the fixed memory state. A probe's measured half-life is the first time its radius reaches 0.04

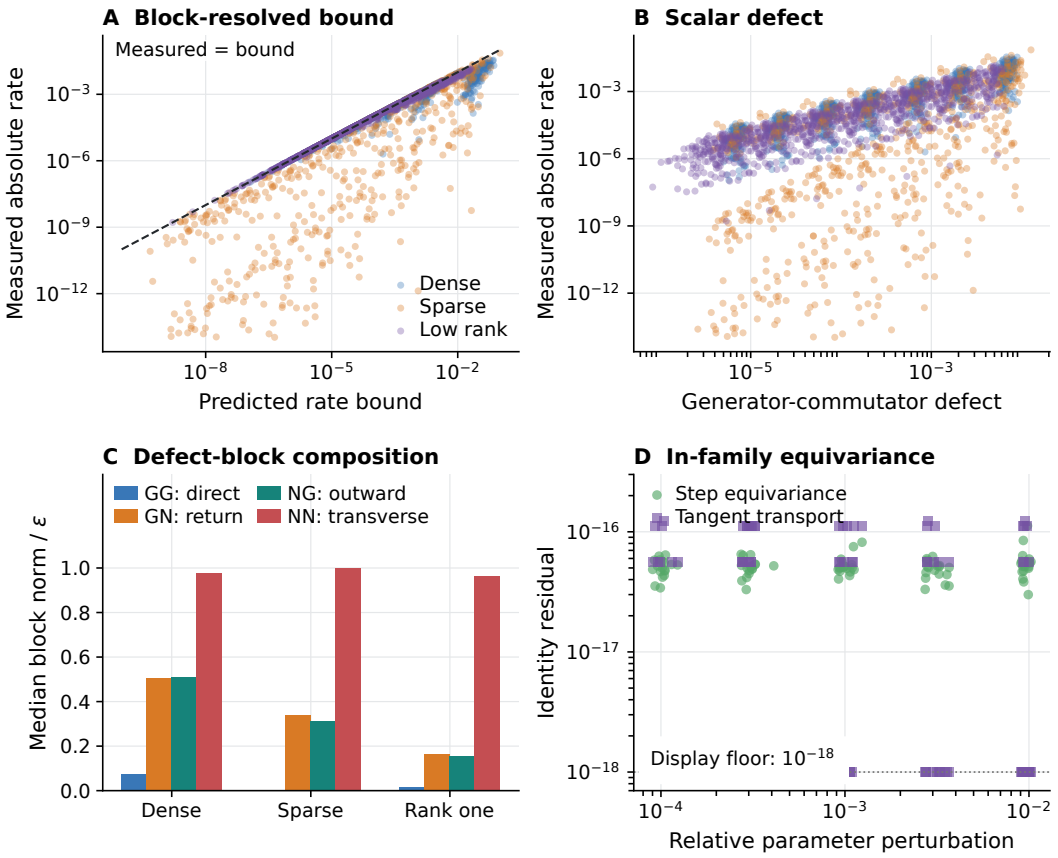


Figure 7: **Parameter and implementation-defect analysis.** (A) Rates of valid ambient-residual cases remain below the dashed bound-equality line. (B) A scalar generator-commutator defect does not identify the route. (C) Median induced block norms divided by perturbation scale; GG is direct tangent damage, NG outward coupling, GN return coupling, and NN transverse damage. (D) Native weight noise inside the exact parameterization preserves step equivariance and analytical tangent transport to numerical precision; values below 10^{-18} are drawn at that display floor, not recorded as positive residuals of that size. This is a local residual-family analysis, not a learned or hardware-specific defect distribution.

or below. If a probe does not reach this threshold within the finite-time dynamics horizon, its exact half-life is unknown and the case is treated separately. The reported case value is the median crossing time across probes.

H Decoder-coordinate stochastic geometry

H.1 Itô covariance and the isotropic noise floor

Let the ambient dynamics have local Itô form (Øksendal, 2003)

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \Sigma(X_t)\,\mathrm{d}W_t, \qquad Q = \Sigma\Sigma^\top. \tag{46}$$

For a C^2 local coordinate ψ, Itô's formula gives

$$\mathrm{d}\psi(X_t) = \left[D\psi\, b + \frac{1}{2}\,\mathrm{Hess}(\psi) : Q\right]\mathrm{d}t + D\psi\,\Sigma\,\mathrm{d}W_t.$$

The Hessian term contributes drift. The instantaneous quadratic variation is exactly

$$\mathrm{d}[\psi]_t = ZQZ^\top\,\mathrm{d}t, \qquad Z = D\psi. \tag{47}$$

Table 7: Summary of the two-coordinate multi-step evaluation. Coverage is evaluated only for cases satisfying the sufficient conditions. Large-scale cases that fall outside these conditions are retained as boundary controls.

Diagnostic	Result
Fixed checkpoints / protected dimension	6 / $q = 2$
Cases through $\epsilon = 10^{-2}$	$90/90$ satisfy conditions
Boundary cases at $\epsilon = 3 \times 10^{-2}$	$0/18$ satisfy conditions
Exponent / graph-radius / angle coverage	$100\%/100\%/100\%$
Bound–measured local displacement correlation	0.9911
Monotone global endpoint error units	$18/18$
Fully uncensored seed–adapter–scale cases	41 (16 probes each)
Primary log correlation; median ratio	0.9996; 1.0119
Finite-median sensitivity cases	56 (15 partially censored)
Sensitivity log correlation; median ratio	0.9788; 1.0055

Proof of Theorem 6. The constraint $Z\bar{A} = I$ implies $Z = \bar{A}^\dagger + R$ with $R\bar{A} = 0$. Since $(\bar{A}^\dagger)^\top = \bar{A}(\bar{A}^\top\bar{A})^{-1}$, $R(\bar{A}^\dagger)^\top = 0$, and therefore

$$ZZ^\top = (\bar{A}^\top\bar{A})^{-1} + RR^\top. \tag{48}$$

Multiplication by σ^2 proves the positive-semidefinite lower bound. The assumption $\sigma > 0$ is needed for the equality characterization: at $\sigma = 0$ both covariance matrices vanish for every faithful decoder. The inequalities then hold trivially, but equality identifies neither a unique decoder nor an optimal action geometry. For $\sigma > 0$, equality holds exactly when $RR^\top = 0$, equivalently $R = 0$ and $Z = \bar{A}^\dagger$. □

The condition $Z = \bar{A}^\dagger$ means the coordinate differential is the minimum-Euclidean-norm left inverse of the action and annihilates the orthogonal normal complement. A decoder can remain perfectly faithful along the orbit ($Z\bar{A} = I$) while having additional normal sensitivity; that sensitivity adds the positive-semidefinite term $\sigma^2 RR^\top$.

H.2 Discrete update bridge

Corollary 5 (One-step discrete covariance)**.** *Let $X_{k+1} = F(X_k) + \eta_{\Delta t}$, set $y = F(x)$, and condition on $X_k = x$. If $\mathbb{E}\eta_{\Delta t} = 0$, $\mathrm{Cov}(\eta_{\Delta t}) = Q_y\Delta t$, ψ is differentiable at y, and its Taylor remainder satisfies*

$$\mathbb{E}\|\psi(y + \eta_{\Delta t}) - \psi(y) - Z_y\eta_{\Delta t}\|^2 = O(\Delta t^2),$$

then

$$\mathrm{Cov}[\psi(X_{k+1}) \mid X_k = x] = Z_y Q_y Z_y^\top \Delta t + o(\Delta t). \tag{16}$$

Proof of Corollary 5. Taylor expansion at $y = F(x)$ gives

$$\psi(y + \eta_{\Delta t}) = \psi(y) + Z_y\eta_{\Delta t} + r_{\Delta t},$$

with $\mathbb{E}\|r_{\Delta t}\|^2 = O(\Delta t^2)$ by assumption, whereas $\mathbb{E}\|Z_y\eta_{\Delta t}\|^2 = O(\Delta t)$. Hence the covariance of the linear term is $Z_y Q_y Z_y^\top \Delta t$, the covariance of the remainder is $O(\Delta t^2)$, and their cross covariance is $O(\Delta t^{3/2})$ by Cauchy–Schwarz. Centering changes only these remainder orders, proving Equation (16). □

A bounded Hessian on the increment support together with $\mathbb{E}\|\eta_{\Delta t}\|^4 = O(\Delta t^2)$ is one sufficient condition for the remainder assumption. For Gaussian increments and a smooth decoder around a state separated from its coordinate singularities, the same local conclusion follows after controlling the vanishing tail outside that neighborhood.

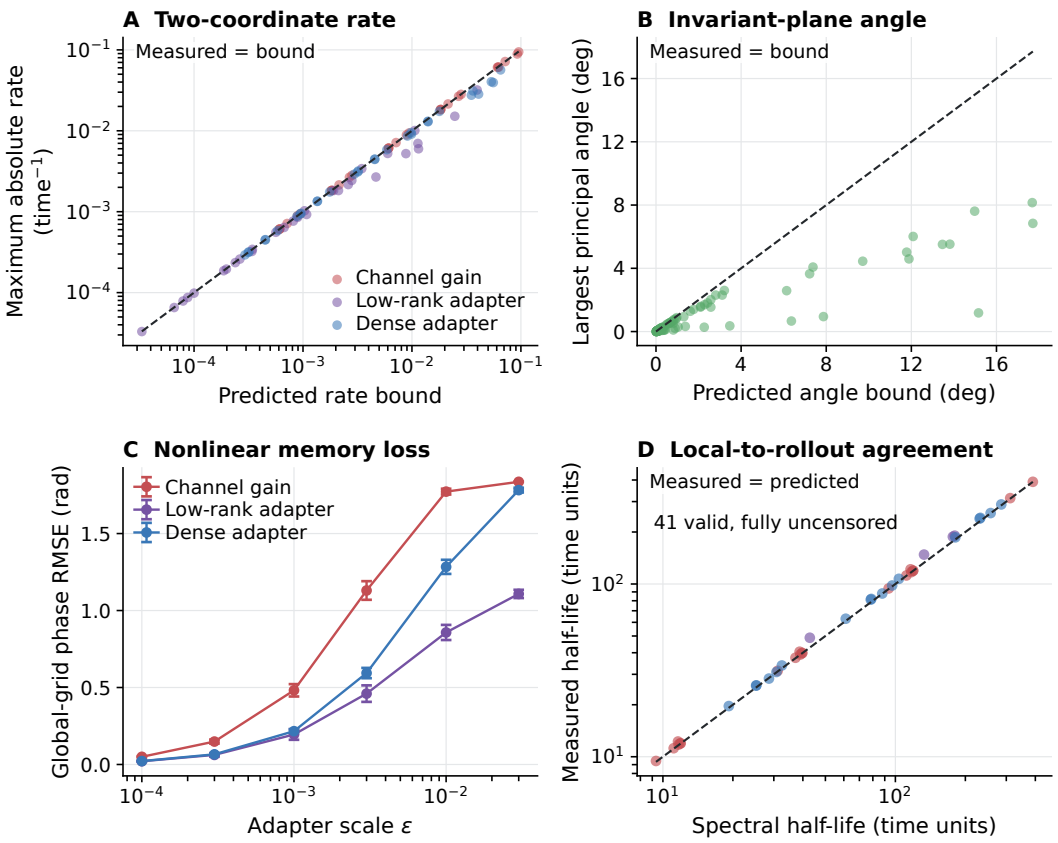


Figure 8: **Local predictions and multi-step dynamics in a two-coordinate memory.** (A,B) Cases satisfying the sufficient conditions remain within the predicted rate and invariant-plane angle bounds; dashed lines mark equality between the measured quantity and its bound. The 18 largest-scale cases fall outside the sufficient regime and are omitted from A/B but retained in the tables. (C) Endpoint phase RMSE after multi-step evolution from the global phase grid, reported as mean ± SEM over six base checkpoints for each adapter and perturbation scale. Error increases monotonically with perturbation scale for all 18 seed–adapter combinations. (D) Median measured half-life of nearby perturbations versus the local two-mode spectral half-life, both in physical-time units, for the 41 cases in which all 16 probes reach the half-radius threshold within the finite-time dynamics horizon. The dashed line marks prediction–measurement agreement rather than a bound. Cases containing probes that do not reach the threshold within the finite-time dynamics horizon are analyzed separately. Panels C and D compare local predictions with subsequent multi-step dynamics; they are diagnostic evidence, not consequences of a global nonlinear theorem.

H.3 Anisotropic ambient noise

Proposition 13 (Generalized least-squares decoder)**.** *Let $Q \succ 0$. Among all Z satisfying $Z\bar{A} = I$, the unique decoder minimizing $ZQZ^\top$ in Loewner order is*

$$Z_* = (\bar{A}^\top Q^{-1} \bar{A})^{-1} \bar{A}^\top Q^{-1},$$

and the minimum covariance is $(\bar{A}^\top Q^{-1} \bar{A})^{-1}$.

Proof. Set $Y = ZQ^{1/2}$ and $\widetilde{A} = Q^{-1/2}\bar{A}$. The constraint becomes $Y\widetilde{A} = I$. Every left inverse has $Y = \widetilde{A}^\dagger + R$ with $R\widetilde{A} = 0$, so $YY^\top = (\widetilde{A}^\top \widetilde{A})^{-1} + RR^\top$. Transforming back gives the stated Z_* and covariance. □

This is the generalized least-squares solution of Aitken (1936); the new use here is as a local recurrent-memory decoder, not as a new linear-estimation identity.

This formula clarifies the scope of the Euclidean tight-frame result. For a fixed action $\bar{A}$ and structured noise Q, Equation (18) optimizes the decoder only. If the action geometry is also a design variable under the whitened-energy constraint $\mathrm{tr}(\bar{A}^\top Q^{-1}\bar{A}) = P$, applying the fixed-energy argument to $Q^{-1/2}\bar{A}$ makes the optimum isotropic in that whitened metric. A fixed raw Euclidean energy budget is a different problem and need not have the same optimizer.

H.4 Proof of the fixed-energy corollary

Proof of Corollary 7. Let $\mu_1, \dots, \mu_q > 0$ be the eigenvalues of $\bar{A}^\top\bar{A}$ with $\sum_i \mu_i = P$. By Theorem 6, every faithful decoder satisfies $D_\psi \succeq \sigma^2(\bar{A}^\top\bar{A})^{-1}$. Hence

$$\lambda_{\max}(D_\psi) \geq \frac{\sigma^2}{\min_i \mu_i} \geq \frac{\sigma^2 q}{P},$$

and the harmonic–arithmetic mean inequality gives

$$\mathrm{tr}(D_\psi) \geq \sigma^2 \sum_i \mu_i^{-1} \geq \frac{\sigma^2 q^2}{P}.$$

Equality in either scalar bound requires all $\mu_i = P/q$. Equality in the matrix lower bound additionally requires $Z = \bar{A}^\dagger$. Thus both bounds are attained exactly under the two conditions stated in the corollary. □

When the rows of $\bar{A}$ are viewed as frame vectors for the memory-coordinate space, $\bar{A}^\top\bar{A} = (P/q)I$ is the finite tight-frame condition. Equivalently, $\bar{A}$ is a scaled isometric embedding from quotient coordinates into the state tangent space. The inequality itself is standard frame/convex algebra; the paper's claim is the decoder-aware recurrent-memory interpretation and causal test (Benedetto & Fickus, 2003; Xu et al., 2023; 2025).

H.5 Angle decoder for the T^2 experiment

For one circular block $z = r(\cos\theta, \sin\theta)$, the group tangent is $a = (-z_2, z_1)^\top$ and the local angle differential is

$$D\theta = \frac{1}{r^2}(-z_2, z_1) = a^\dagger.$$

The T^2 action is free on the target orbit, so the natural generator basis gives $\bar{A} = A$. For two disjoint circular blocks, $\bar{A}$ is block diagonal in its two columns and the blockwise angle decoder satisfies $D\psi = \bar{A}^\dagger$ exactly on the target orbit. Consequently, the prescribed-geometry study tests the equality case of Theorem 6, not merely its lower bound.

I Prescribed fixed-budget representation geometry

I.1 Prescribed representation and training settings

For target condition number $\kappa \geq 1$ and tangent energy $P = 2$, the two circle radii are

$$r_{\min} = \sqrt{\frac{2}{1+\kappa^2}}, \qquad r_{\max} = \kappa r_{\min}. \tag{49}$$

At any target-orbit state, the two singular values of the T^2 action matrix are precisely $r_{\min}$ and $r_{\max}$. The architecture enforces independent rotations of the two state blocks and uses only invariant features in learned coefficients. Each condition is trained on the same two-angle velocity-input path-integration distribution, optimizer family, update budget, and architecture size. Seed j at one geometry is paired with seed j at every other geometry.

These conditions are prescribed, scale-conjugate representations with matched noiseless input–output dynamics by construction: the angle decoder removes the radius scaling, while training uses the same deterministic task in each condition. This paired design isolates the effect of isotropic ambient state noise; it is not a sample of independently discovered representation geometries.

The experiment pairs six seed identities across six prescribed conditions, producing 36 evaluation cells. Summaries use the six seed identities or condition means as specified below, not 36 independent model replicates. Geometry is prescribed by the radii. Training learns task dynamics inside that representation and does not test spontaneous tight-frame discovery.

Table 8: Condition-level results at test horizon 256 and speed scale 1.5. Uncertainty is SEM over the six paired seeds.

$\kappa(\bar{A})$	Predicted $\lambda_{\max}$	Long-horizon simulation $\lambda_{\max}$	Noisy RMSE	Clean RMSE
1.0	0.000900	0.000920 ± 0.000014	0.204 ± 0.014	0.135 ± 0.021
1.5	0.001462	0.001481 ± 0.000015	0.225 ± 0.013	0.134 ± 0.020
2.0	0.002250	0.002337 ± 0.000047	0.266 ± 0.012	0.136 ± 0.021
3.0	0.004500	0.004490 ± 0.000058	0.353 ± 0.010	0.136 ± 0.022
5.0	0.011700	0.011772 ± 0.000126	0.543 ± 0.007	0.136 ± 0.020
8.0	0.029250	0.030317 ± 0.000489	0.843 ± 0.010	0.134 ± 0.019

The maximum absolute tangent-energy deviation from two is 2.519×10^{-7}. Maximum double-precision step-equivariance and group-tangent covariance residuals are 1.999×10^{-16} and 2.483×10^{-16}. The median relative errors of the full local and long-horizon simulation diffusion tensors are 0.560% and 3.317%. Pearson correlation between predicted and long-horizon simulation worst diffusion is 0.998912 across all 36 cells and 0.999920 across the six condition means; condition-mean Spearman correlation is one.

Every one of the six seed identities has strictly increasing noisy horizon-256 worst-coordinate RMSE. The paired endpoint difference $\mathrm{RMSE}_{\kappa=8} - \mathrm{RMSE}_{\kappa=1}$ is 0.638943 ± 0.010973 SEM and is positive for all six seeds. In contrast, the range of clean condition means is 0.002875, the seed-fixed clean slope is statistically indistinguishable from zero in the regression specified before analysis, and the maximum validation-loss range across geometries within a seed is 3.492×10^{-10}. Decoder sensitivity grows from 1.0 to 5.70, as expected from the shrinking radius of the weakest coordinate.

Noise is added after each recurrent update as $\eta_k \sim \mathcal{N}(0, \sigma^2 \Delta t I)$ with $\Delta t = 0.1$. Therefore Corollary 5 is the formal one-step, small-Δt bridge. The reported finite-step tensor fit and horizon-256 task curves additionally test repeated nonlinear composition and should not be read as an exact continuous-time covariance identity at $\Delta t = 0.1$.

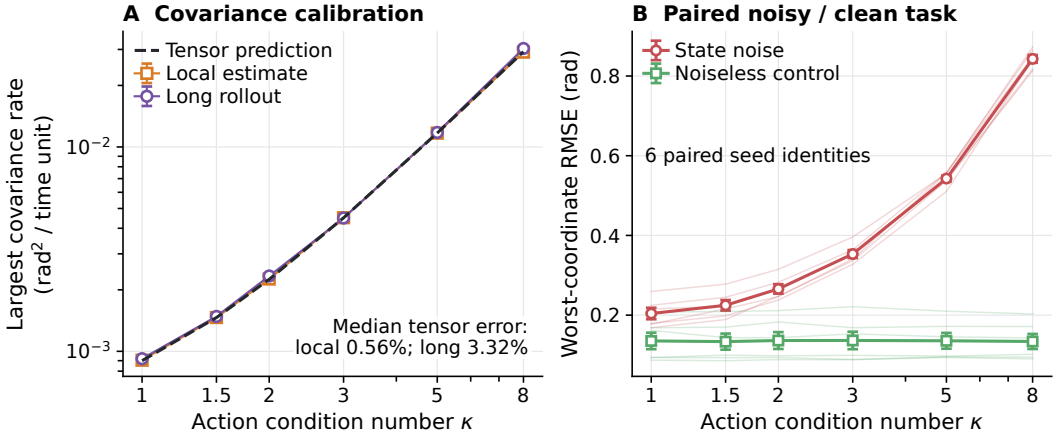


Figure 9: **Prescribed-geometry calibration underlying the learned-geometry study.** (A) The dashed least-squares prediction, hollow local-estimate squares, and hollow long-horizon simulation circles show worst-coordinate covariance rates. Error bars are mean ± SEM over six paired seed identities; median full-tensor relative errors are 0.56% locally and 3.32% over long-horizon simulations. (B) Thin lines pair each seed across geometries; every noisy curve worsens with anisotropy while the scale-conjugate noiseless control remains nearly flat. Geometry is prescribed here; learned selection is evaluated in Section K.

I.2 SHARED-INPUT TRAJECTORY DIAGNOSTIC

Figure 1 uses a separately logged, matched-input evaluation of fixed checkpoints from the prescribed-geometry study, not additional training or an additional independent experiment. Before diagnostic evaluation, the evaluation design specified $\kappa \in \{1, 8\}$, all six archived seed pairs, and seed 0 (the lowest identity) for the main illustration. Stimulus index 666 is the first correlated-velocity sequence in a 1,000-sample batch from the original sampler with generator seed 2026091401, horizon 256, speed scale 1.5, and $\Delta t = 0.1$. This input was chosen for its temporal structure before viewing model responses and is identical across pairs and geometries. Unlike the original multi-input evaluator, whose generator seed depends on geometry, this diagnostic explicitly matches the input realization.

For each base seed j, 64 repeated noise draws use generator seeds $2026091500 + 1000j + k$, $k = 0, \ldots, 63$. Each paired realization receives the same unscaled full-state Gaussian increments, with covariance $0.03^2 \Delta t I_{12}$ after each unchanged update. The figure displays the preselected trials $k = 0, \ldots, 7$. It retains the actual noiseless paths, with no smoothing, recentering, or trajectory-dependent selection. Angles are decoded by `atan2`, then temporally unwrapped using principal increments and anchored at their initial values. Both panels share their axis limits and equal geometric scaling. The target is retained in the records but not drawn; clean model agreement is not zero task error.

In the displayed seed-0 pair, the complete 64-draw endpoint comparison has larger first-angle variability and smaller second-angle variability under unequal scales. These are descriptive, one-input observations, not independent trained-model replicates or replacements for the original multi-input statistics. They do not imply a universal preference for equal scales or an exact long-horizon covariance law. All 12 clean and 768 noisy paths, input and noise arrays, checkpoint identities, selection procedure, and the six-pair display are retained in the supplementary `clean_noisy_memory/` package.

J THEORY-GUIDED LEARNING OF ACTION GEOMETRY

J.1 EXACT PARAMETERIZATION AND TRAINING CONDITIONS

The learned-geometry study retains the prescribed-geometry recurrent cell, blockwise angle decoder, two-angle velocity task, $P = 2$ budget, $\sigma = 0.03$ ambient noise, $\Delta t = 0.1$, 180-update training budget, and horizon-256 evaluation. Only the orbit radii become trainable. Two logits enforce

$$w = \mathrm{softmax}(\alpha), \qquad r_i^2 = 2w_i, \qquad r_1^2 + r_2^2 = 2 \tag{50}$$

identically up to floating-point roundoff. The group action remains independent rotation of the two state blocks, and every learned coefficient depends only on invariant features. Thus geometry learning preserves exact T^2 equivariance rather than approximating it through augmentation.

Five paired training conditions use the same recurrent initialization and minibatch stream for a given seed: fixed $\kappa = 8$, learnable geometry with clean loss, learnable geometry with state noise, clean loss plus the normalized covariance penalty

$$\mathcal{L}_{\mathrm{geom}} = \tfrac{1}{2} \operatorname{tr}\big[(\bar{A}^\top \bar{A})^{-1}\big] - 1, \tag{51}$$

and a fixed $\kappa = 1$ oracle. All training conditions contain the same total number of parameters; fixed training conditions simply disable gradients for the two geometry logits. The clean scale-conjugacy test verifies that changing fixed radii does not add deterministic task capacity.

J.2 HYPERPARAMETER SELECTION AND EVALUATION DESIGN

The covariance-regularization coefficient was selected using a clean-only preliminary run with seeds 0 and 1, 120 updates, and candidate coefficients $\{10^{-3}, 3 \times 10^{-3}, 10^{-2}\}$. The initial conditioning target $\kappa \le 1.5$ was not reached and was revised to $\kappa \le 2.5$ before evaluating the original six-unit study. The smallest coefficient meeting the revised target while preserving matched clean validation performance, 10^{-3}, was selected. No noisy test metric entered this decision. Additional development evaluations at coefficient 3×10^{-3} with seed labels 0–5 were excluded from the original evaluation summaries.

The original evaluation used labels 6–11, which were not used for model or hyperparameter selection. Subsequent inspection found that its arithmetic generator-seed formulas could reuse random streams across training roles and adjacent units. Distinct displayed labels therefore did not establish independent random streams. We retain these observations as descriptive historical records, do not interpret their nominal SEMs as uncertainty from independent training replicates, and do not pool them with the independent evaluation set. The full details, threshold change, excluded development records, six-unit results, and structural diagnostics are retained in the supplementary material; `historical_t2/README.md` identifies the archived records and their interpretation.

Our principal T^2 learning evidence is the independent twenty-unit replication in Section K. Its settings were fixed after inspecting the original results and before evaluating the new evaluation set, with the selected coefficient and training settings held fixed. The revised generator schedule separates random roles and training units while preserving paired initialization and data streams across training conditions within each unit. The replication's outcomes, uncertainty estimates, and clean-error tradeoffs are reported in Section K.

K Independent replication and retrospective bound comparison

K.1 Replication design and statistical units

After inspecting the original results, we specified the replication settings and recorded the code version before evaluating the new evaluation set; the study was not externally preregistered. Twenty training units, labeled $10000, 10100, \ldots, 11900$, each trained all five training conditions without selecting or replacing seeds. The architecture, loss, coefficient 10^{-3}, 180 updates, batch size 96, training horizon 64, invariant hidden size 8, feature width 24, learning rates 0.003 for recurrent parameters and 0.02 for geometry, $\Delta t = 0.1$, $P = 2$, and $\sigma = 0.03$ were held fixed. The same initialization and data streams were shared across training conditions within each unit, not across units. All 100 runs in the final evaluation completed with finite recorded metrics.

To avoid reuse of random streams, we assigned independent deterministic seeds to each study dimension, training unit, stochastic role, and evaluation index using a SHA-256-based construction. We verified that no assigned seeds collided. Training data, training noise, validation data, test data, test noise, and bootstrap draws have distinct roles; preliminary run identities are separate. An implementation regression mapped the substituted seeds back to the archived values and recovered identical training curves and weights. The replication used the same data-generating distribution with corrected random-stream separation; it was not intended to reproduce the original run exactly.

The noisy task uses 1,000 trajectories per model at horizon 256 and speed scale 1.5. For each coordinate, circular RMSE is computed across the endpoint errors of the evaluation trajectories; the reported metric is the maximum of these coordinatewise RMSEs. Local covariance uses 40,000 increments per model; long covariance uses 1,500 zero-input trajectories over 300 updates, with covariance traces and raw increments retained. All metrics were independently reconstructed from those arrays. Trajectories are nested samples, not additional training seeds. The primary contrast was covariance-regularized minus clean-only noisy RMSE; all remaining contrasts are descriptive secondary analyses.

K.2 Results and clean-error equivalence

The primary paired mean is -0.622977 rad, with SD 0.015187, SEM 0.003396, a two-sided 95% t_{19} interval $[-0.630085, -0.615869]$, and a 10,000-resample seed-bootstrap interval $[-0.629221, -0.616291]$. All twenty differences are negative. The covariance-regularized condition's paired clean change is -4.36×10^{-8} rad, with interval $[-1.28 \times 10^{-7}, 4.09 \times 10^{-8}]$, contained in the prespecified $[-0.01, 0.01]$ rad equivalence margin. This establishes equivalence under that operational margin, not mathematical equality of trained predictors.

Direct noise training has paired noisy change -0.620221 rad but clean change $+0.018431$ rad, with 95% interval $[0.006266, 0.030595]$. It does not meet the clean-equivalence criterion. Thus the original six-unit clean-error observation does not extend to this larger evaluation set for the noise-trained condition. No adverse run is discarded. Covariance regularization still trails the fixed isotropic oracle in noisy RMSE by 0.050285 ± 0.001434 rad. The original six-unit observations and the descriptive

Table 9: Independent replication with twenty paired training units per training condition. Values are mean ± SEM across training units. RMSE is in radians; κ is the action-map singular-value condition number.

Training Condition	Final κ	Noisy RMSE	Clean RMSE
Fixed anisotropic	8.000 ± 0.000	0.845 ± 0.003	0.103 ± 0.005
Learnable clean	8.065 ± 0.016	0.852 ± 0.004	0.103 ± 0.005
Learnable noisy	1.757 ± 0.005	0.232 ± 0.003	0.121 ± 0.006
Covariance regularized	1.755 ± 0.000	0.229 ± 0.002	0.103 ± 0.005
Fixed isotropic oracle	1.000 ± 0.000	0.179 ± 0.003	0.103 ± 0.005

comparison with the original six-unit study remain available separately in the supplementary material; they are not pooled with this replication. Full SDs, SEMs, paired intervals, and all seed-level values are provided in the supplementary material, together with analyses that retain unsuccessful runs.

K.3 Retrospective usefulness of resolving defect blocks

We reconstruct 4,566 archived maps or finite sequences in their original reducing frames, verifying their four block norms against the saved tables. In those same coordinates we compute the true full operator norm $\delta = \|E\|_2$, rather than substituting the intervention amplitude or a 2×2 norm envelope. Since every block norm is at most δ, a structure-blind comparator can validly pass $a = b = c = e = \delta$ to the same sufficient-condition algorithm. Both methods use identical normal contraction, durations, endpoint conditioning, and specified initial graph. This is one norm-only comparator, not the best possible norm-only theorem.

Table 10: Retrospective bound availability, with all archived cases retained. The denominator is the complete case set, including failed sufficient conditions. These nested cases are not independent training seeds. Neither method has a measured coverage violation among its valid cases.

Archived study	Cases	Block-resolved valid	Norm-only valid
Designed routing	1,608	1,584	1,404
Ambient residuals	2,520	2,515	2,160
Native residual coordinates	90	90	90
Learned T^2 adapters	108	90	72
Finite-horizon sequences	240	216	216

Invalid bounds remain unavailable, never zero. Sharpness ratios are computed with all zero-defect controls retained: the 1,584 valid designed-routing cases include 12 zero-defect cases, whereas 1,572 counts only nonzero defects. Ratios are computed only on the common valid subset and above the original numerical floors and duration-adjusted coverage tolerance. The finite-horizon comparison retains all 24 out-of-regime controls and does not show an availability advantage. Analytic three-dimensional controls with identical block norms give positive, negative, or zero center displacement according to leakage orientation; a neutral control can also fail the sufficient criterion. Norm information therefore does not identify the sign of displacement, and failure of the sufficient criterion is not evidence of instability. This retrospective analysis neither extends the theorem's scope nor provides an online guarantee.

K.4 Noise-dependent geometry in coupled torus integrators

Fixed-Euclidean-budget specialization. For orthogonal charge-one planes, write $p_i = r_i^2 > 0$ and fix $\sum_i p_i = P > 0$. The decoder is the local blockwise angle map, not a learned or freely optimized decoder. For fixed ambient block covariance $Q_i = \sigma^2 v_i I_2$, with $v_i > 0$ and $\sigma > 0$, its differential has norm $1/r_i$ on plane i, so Equation (15) gives

$$D_\psi = \operatorname{diag}(\sigma^2 v_i / p_i), \qquad \operatorname{tr} D_\psi = \sigma^2 \sum_i v_i / p_i.$$

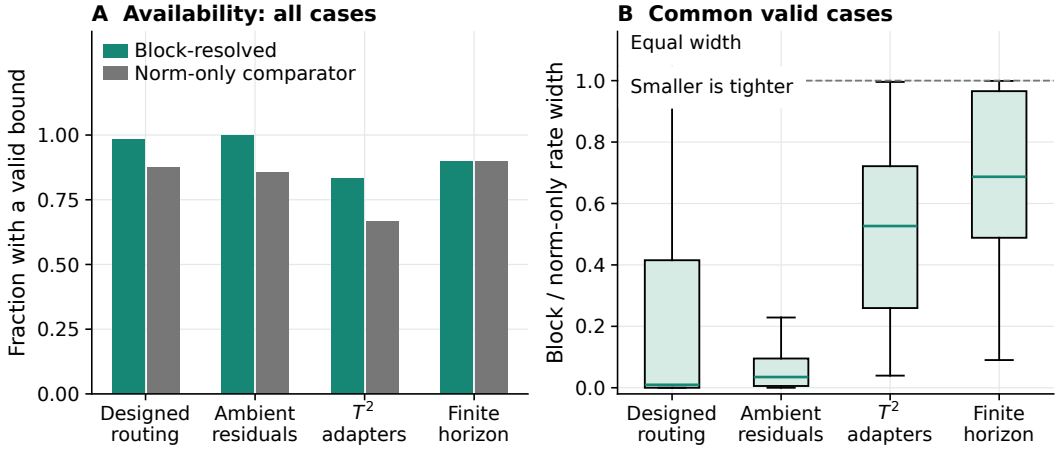


Figure 10: **Retrospective usefulness, not prospective forecasting.** (A) Sufficient-bound availability over all cases in four archived studies. (B) Rate-width ratios on the common valid subset: one means equal width, and smaller means tighter block-resolved intervals. Boxes show quartiles, with 1.5-interquartile-range whiskers; outlier markers are omitted only from the display. All cases remain in the tables; the native-coordinate control is in Table 10. Availability improves in three displayed local studies, not in the finite-horizon study; sharpness and availability are distinct benefits.

Cauchy–Schwarz applied to $\sqrt{v_i/p_i}$ and $\sqrt{p_i}$ gives

$$\left(\sum_i \sqrt{v_i}\right)^2 \le \left(\sum_i v_i/p_i\right)\left(\sum_i p_i\right).$$

Equality holds exactly when p_i is proportional to $\sqrt{v_i}$. Consequently the unique positive trace-minimizing allocation is

$$p_i^* = \frac{P\sqrt{v_i}}{\sum_j \sqrt{v_j}}, \qquad R_v(p) = \frac{P}{(\sum_j \sqrt{v_j})^2}\sum_i \frac{v_i}{p_i} - 1 \ge 0.$$

It is squared radius, not radius, that scales as the square root of the variance multiplier; radius scales as $v_i^{1/4}$. For the different objective $\min \max_i \sigma^2 v_i/p_i$, if $M = \max_i v_i/p_i$, then $P \ge \sum_i v_i/M$, with equality exactly at $p_i = Pv_i/\sum_j v_j$. Equal v_i recover $p_i = P/q$; unequal v_i generally select anisotropic Euclidean geometry. This elementary specialization does not optimize arbitrary full covariance, a global integrable generalized least-squares decoder, or long-time nonlinear task loss. The Euclidean budget here differs from a noise-whitened energy constraint. Invariant hidden-state noise has zero instantaneous derivative under angle decoding but can affect later recurrent dynamics. Numerical checks of R_v and its logit gradient agree with the closed form and central differences. Formalization scope is summarized in Section N.1.

Coupled equivariant architecture. The state consists of q charge-one planes $z_i \in \mathbb{R}^2$ and invariant $h \in \mathbb{R}^{32}$; T^q rotates each z_i independently. Squared target radii are $p_i = q\,\mathrm{softmax}(\alpha)_i$, and $e_i = \|z_i\|^2/p_i - 1$. A shared 64-wide tanh network maps the invariants (e, h, u, u^2) to scalar coefficients c_i, angular coefficients b_i, and hidden coefficients g. With $K = I + 0.2\mathbf{1}\mathbf{1}^\top/q$ and learned positive gains a_i, ω_i, γ, the field is

$$\dot{z}_i = \{-a_i(Ke)_i + 0.04e_i\tanh(c_i)\}z_i + u_i(\omega_i + 0.04b_i)Jz_i,$$
$$\dot{h} = -\gamma h + 0.04\Big(\sum_i e_i + \sum_i u_i^2\Big)g.$$

Every coefficient is invariant, so the field and Euler update are exactly equivariant in real arithmetic. The zero-input target torus with $h = 0$ is fixed and has q independent group tangents; it has no additional nonzero flow vector. The vector-output loss uses $z_i/(\|z_i\| + 10^{-8})$; this readout is not a normalization of the hidden dynamics.

Preliminary runs. Before running the full higher-dimensional evaluation, we tested the original unbounded-radial implementation in four preliminary runs: clean and noisy runs for both T^4 and T^8. The two noisy runs became nonfinite at updates 160 and 188, respectively, so the full evaluation was not started; these failed runs remain archived.

We then made one documented modification, called BR1 in the supplementary materials, replacing $0.04e_ic_i$ with $0.04e_i\tanh(c_i)$ while leaving the rest of the executable model unchanged. The scalar correction is bounded by $0.04|e_i|$, and its vector contribution by $0.04|e_i|\|z_i\|$. These bounds do not provide a global bound on the state or a stability guarantee for the Euler dynamics under unbounded Gaussian noise.

The modified implementation was rerun using the same four model and random-seed identities rather than newly selected ones. All four runs satisfied the predefined numerical stability criteria. We fixed the modified model and evaluation settings before the full evaluation. These settings were documented before the results were examined, although the study was not externally preregistered. The initial failures and the same-identity reruns are retained in the supplementary records.

Evaluation design and pairing. The final evaluation contains 192 runs: six training conditions times sixteen independent paired training units in each of $q = 4, 8$. Training conditions are fixed anisotropic geometry, clean geometry learning, noise-trained geometry, anisotropic-covariance regularization, isotropic regularization, and a fixed analytical trace oracle. The subtraction "anisotropic-covariance minus isotropic regularization" is held fixed when switching evaluation noise; the labels do not imply that one training condition matches every noise condition. Initial nongeometry weights are paired across training conditions. Independent random-number streams are used for different dimensions, runs, and stochastic roles, with collision checks reported in the supplement. There is no pooling with the earlier T^2 studies.

All runs use 360 AdamW updates, horizon 64, batch 96, hidden size 32, feature width 64, $\Delta t = 0.1$, gradient clipping at 1, base learning rate 0.003 and geometry learning rate 0.02. Base weight decay is 10^{-5}; geometry weight decay is zero. The regularizer weight is 0.001. Clean vector-output MSE trains all training conditions except noise training, which injects the declared state noise. Final weights at update 360 are evaluated without best-checkpoint selection. Total parameter counts are 5,493 for T^4 and 6,793 for T^8; the fixed training conditions freeze q geometry parameters. Initial action condition number is 8. Gaussian, piecewise-constant, and correlated velocities use training speed scale 0.8 and uniform initial phases.

Noise has $\sigma = 0.03$ and a seed-permuted geometric profile from 1 to 64, normalized to mean one. The ambient covariance is fixed independently of learned radii: $\sigma^2\,\mathrm{diag}(v_1I_2, \ldots, v_qI_2, I_{32})$. Isotropic evaluation replaces v by ones, with unchanged total covariance trace. The energy budget $P = q$ is fixed within each dimension; it holds per-coordinate energy, not total energy across dimensions, constant. The primary endpoint is mean squared wrapped endpoint phase error over 1,000 trajectories and q coordinates at horizon 256, physical time 25.6, and speed 1.5. It is MSE in rad^2, not time-averaged RMSE or worst-coordinate RMSE. Horizons 64/128 and all other contrasts are secondary. Inference uses sixteen paired training units, not trajectories as replicates. Both paired t intervals and 10,000 paired-unit percentile-bootstrap intervals are reported, without multiplicity-controlled claims for secondary comparisons. No failed or unfavorable evaluation run was dropped.

Task results and control comparisons. Tables 11 and 12 give absolute means and paired differences. The T^4 primary advantage and corresponding T^8 secondary advantage reverse under isotropic evaluation noise. All 36 declared clean per-coordinate contrast intervals lie inside ± 0.01 rad, a conditional equivalence criterion for these contrasts, not evidence that every training objective is cost-free. In particular the earlier T^2 noise-training cost $+0.018431$ rad, interval $[0.006266, 0.030595]$, remains an adverse result. The T^4 covariance-minus-oracle anisotropic-noise difference is $+0.000848168$ rad^2: its t interval crosses zero while its bootstrap interval does not. Neither demonstrates equivalence. The T^8 difference $+0.000280051$ has both intervals above zero. The local trace oracle is not claimed optimal for this nonlinear finite-horizon task.

Geometry swaps change only the final clean-trained cell's geometry logits; nongeometry weights, inputs, and noise streams are identical within each comparison. Swapping to trace-optimal geometry lowers anisotropic-noise MSE in all sixteen units per dimension. Swapping to isotropic geometry improves the mean but worsens one T^4 unit under anisotropic noise (15/16 improve). Clean MSE

Table 11: Absolute endpoint phase MSE in rad^2 at horizon 256, speed 1.5. Mean $\pm$ SEM across sixteen training units per training condition and dimension; clean, anisotropic-noise, and isotropic-noise evaluations use the same final cells.

q	Training condition	Clean	Anisotropic	Isotropic
4	Fixed anisotropic	0.002382 ± 0.000031	0.134828 ± 0.030448	0.157644 ± 0.001403
4	Clean learning	0.002382 ± 0.000031	0.134851 ± 0.030542	0.157763 ± 0.001327
4	Noise training	0.002724 ± 0.000069	0.017314 ± 0.000467	0.038315 ± 0.001035
4	Anisotropic-cov. loss	0.002382 ± 0.000031	0.017494 ± 0.000461	0.038390 ± 0.001053
4	Isotropic loss	0.002382 ± 0.000031	0.023600 ± 0.000279	0.024161 ± 0.000101
4	Trace oracle	0.002382 ± 0.000031	0.016646 ± 0.000089	0.040753 ± 0.000231
8	Fixed anisotropic	0.002561 ± 0.000032	0.106317 ± 0.014793	0.109176 ± 0.000602
8	Clean learning	0.002561 ± 0.000032	0.106292 ± 0.014806	0.109130 ± 0.000572
8	Noise training	0.002822 ± 0.000040	0.017743 ± 0.000124	0.033753 ± 0.000559
8	Anisotropic-cov. loss	0.002561 ± 0.000032	0.018047 ± 0.000123	0.033974 ± 0.000584
8	Isotropic loss	0.002561 ± 0.000032	0.024234 ± 0.000165	0.024160 ± 0.000106
8	Trace oracle	0.002561 ± 0.000032	0.017767 ± 0.000091	0.036163 ± 0.000151

Table 12: Paired endpoint MSE differences in rad^2. A means anisotropic evaluation noise and I isotropic noise; the covariance-minus-isotropic training condition subtraction never changes. Only the starred row is primary. Other rows are descriptive and the two interval procedures are retained even when they differ.

q	Contrast	Noise	Mean	95% t interval	95% bootstrap
4	Cov.–isotropic*	A	-0.006106	$[-0.006674, -0.005538]$	$[-0.006567, -0.005563]$
8	Cov.–isotropic	A	-0.006188	$[-0.006335, -0.006040]$	$[-0.006323, -0.006061]$
4	Cov.–isotropic	I	$+0.014229$	$[+0.011963, +0.016496]$	$[+0.011913, +0.015956]$
8	Cov.–isotropic	I	$+0.009814$	$[+0.008736, +0.010891]$	$[+0.008847, +0.010767]$
4	Cov.–oracle	A	$+0.000848$	$[-0.000074, +0.001771]$	$[+0.000076, +0.001747]$
8	Cov.–oracle	A	$+0.000280$	$[+0.000091, +0.000469]$	$[+0.000125, +0.000460]$

changes are numerical nulls: absolute paired means are at most 2.1×10^{-10} rad^2. Deterministic phase conjugacy is checked separately; the 10^{-8} output-normalization offset permits a small radius-dependent vector magnitude difference, so vector readouts are not claimed algebraically identical. All swap pairs and both interval estimates are included, not just favorable mean effects.

Local covariance and finite-step sensitivity. Local covariance uses 40,000 increments at each of three phase points with $\Delta t = 5 \times 10^{-5}$; full centered covariance matrices are divided by physical time. Long zero-input covariance uses 1,500 trajectories, 300 steps at $\Delta t = 0.1$, unwrapped phase, and a full-matrix covariance-versus-time linear fit. The local decoder prediction is not itself a long-time theorem. Across cells, three-phase-averaged local relative Frobenius error ranges from 0.27 to 1.50 percent; long-horizon simulation error ranges from 1.14 to 12.43 percent. No range is truncated in Figure 11. The long-horizon simulations record 7,325 branch-wrap events, two phase jumps larger than $\pi/2$, no radii below 10^{-3}, and no nonfinite state values. These observed event counts are not a proof of global stochastic stability.

Six fixed cases were also evaluated at $\Delta t = 0.05$ for 600 steps, keeping physical duration 30. Table 13 retains all six outcomes: tensor error is larger in each half-step evaluation. These runs use separate noise streams, not coupled Brownian refinements; two step sizes therefore do not establish discretization convergence. Relative trace changes range from -1.583% to $+1.606\%$.

Structural checks and corrected operator norms. Final-checkpoint loading, parameter counts, invariant covariance construction, and stored normal-block reconstruction are rechecked for all 196 cells. Across the final evaluation, the maximum float64 step-equivariance and direct tangent-transport errors are below 2.34×10^{-16} and 3.15×10^{-16}, respectively; action singular values remain positive in the sampled checks. These are real-arithmetic architecture identities tested numerically, not finite-precision exactness.

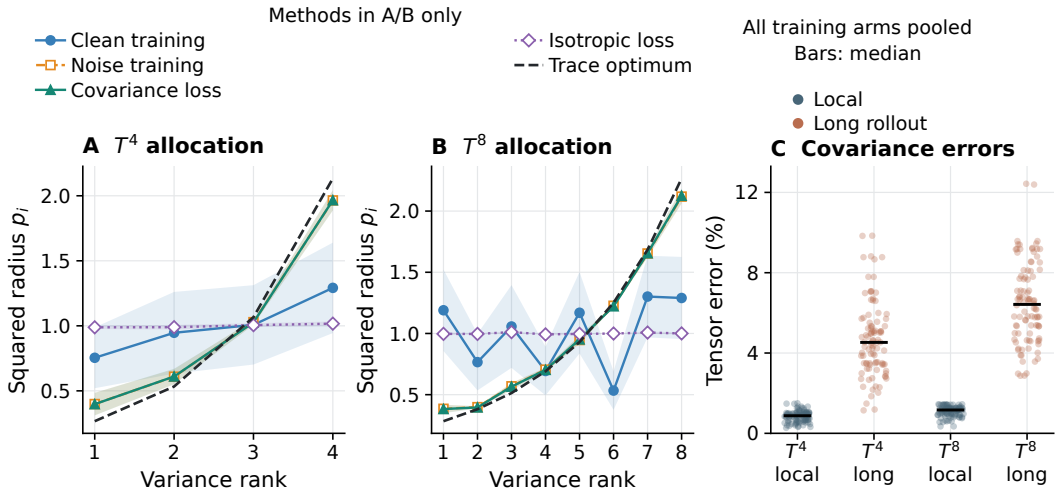


Figure 11: **Noise allocation and covariance calibration.** (A/B) Learned squared radii, aligned within each seed by variance rank before averaging; bands are SEM over sixteen training units. The method legend applies only to A/B. Hollow squares and dashes distinguish noise training from the covariance-loss triangles; the analytic trace optimum and isotropic allocation are shown. (C) Local and long-horizon simulation full-tensor relative Frobenius errors pool all training conditions, using a separate estimator legend; black bars are medians. Each estimator displays all 192 evaluation cells (96 per dimension), without truncating the range. These are not independent covariance replicates; finite-step long-horizon simulation agreement is not an asymptotic stochastic theorem.

Table 13: All fixed step-size sensitivity cases. Errors are relative Frobenius percentages; trace change is half-step minus nominal, divided by nominal. Each row uses the first training unit and independent noise streams at the two step sizes.

q	Training condition	Error at 0.1	Error at 0.05	Trace change (%)
4	Fixed anisotropic	4.631	5.665	−0.602
4	Trace oracle	5.083	6.680	−1.583
4	Anisotropic-cov. loss	5.083	6.680	−1.583
8	Fixed anisotropic	5.802	7.558	+1.606
8	Trace oracle	6.680	8.766	−0.549
8	Anisotropic-cov. loss	6.685	8.761	−0.502

An installed-runtime discrepancy affected the original descriptive normal operator-norm field in 95 of 196 records (92 final-evaluation records and three preliminary runs). The saved matrices are unchanged. NumPy SVD, the square root of the largest Gram eigenvalue, and the explicit maximum of PyTorch singular values agree. A minimal example has original value 0.6650574 but corrected norm 1.0795389; the standalone fixture records shape, strides, dtype, software, and exact call. The cause is unresolved, and no claim about other installations is made. We matched the corrected values to the corresponding records by study stage, dimension, training unit, and condition. Corrected operator norms are at least one in 153/196 cells, including 149/192 evaluation cells.

All recomputed normal spectral radii remain below one, as required by the predefined spectral-radius criterion. Spectral radius below one is not one-step Euclidean contraction, nor does it establish contraction of arbitrary time-varying products or global nonlinear noisy stability. Therefore these checks do not instantiate the normal-contraction hypotheses of the deterministic perturbation propositions. This field was not used in training, checkpoint selection, task evaluation, or covariance estimation. The separate deterministic studies use NumPy operator norms and retain their own normal-contraction requirements. These stochastic results do not extend the deterministic propositions or the Lean formalization.

L ADDITIONAL RELATED WORK

Nonautonomous and subspace-perturbation theory provides the background for our finite-horizon specialization (Sacker & Sell, 1976; Coppel, 1978; Pötzsche, 2010; Wedin, 1972). The result conditions on a supplied recurrent trajectory rather than asserting a general dichotomy theorem. Goldstone-like propagation is studied by Iqbal et al. (2026); our emphasis is on identifiable memory tangents, block-resolved perturbations, and decoder-conditioned covariance.

Recent work studies learned manifolds, approximate persistence, symmetry regularization, and infinite-horizon approximation of normally hyperbolic continuous attractors (Darshan & Rivkind, 2022; Ságodi et al., 2024; Liang et al., 2025; Ságodi & Park, 2026). Connectivity roughness can induce drift toward discrete minima, while constrained optimization can stabilize multiple embedded attractor manifolds (Agmon & Burak, 2023). Robust memory manifolds can also arise without exact symmetry (Can & Krishnamurthy, 2025). Cotteret et al. analyze a complementary stability–resolution tradeoff under noise and heterogeneity for distributed grid-cell-like codes; our stochastic result instead conditions decoded covariance on a fixed action, decoder, metric, and energy budget (Cotteret et al., 2025). Representation-theoretic treatments broaden equivariant architectures (Kondor & Trivedi, 2018), while approximate variants relax hard constraints (Wang et al., 2022). Our bounds instead condition local memory degradation on four measured tangent/normal blocks and give no global guarantee from a scalar defect. Recurrent Equivariant Constraint Modulation learns layerwise relaxation from data-dependent symmetry gaps (Pertigkiozoglou et al., 2026). Our setting is complementary: exact per-input equivariance protects the reference memory bundle, while explicit off-equivariant defects are diagnosed through their tangent/normal blocks rather than represented by learned relaxation levels. Recent hippocampal RNN work reports a learned mixed-symmetry regime in which dominant symmetric recurrence stabilizes an attractor and a weaker antisymmetric component generates directed flow (Wagner et al., 2026). This is complementary to our setting: Wagner et al. study learned motion along an attractor, whereas our bounds condition degradation of a protected memory bundle on measured tangent/normal perturbation blocks.

Certified equivariant world models bound different objects: simulation/rollout error across symmetry configurations, Lyapunov-limited horizons, and conformally calibrated orbit-valid trust regions (Wang, 2026a;b). Our analysis concerns continuation and rotation of the protected group-tangent bundle itself. Numerical Lyapunov algorithms (Oseledets, 1968; Benettin et al., 1980b;a) and covariant subspaces (Ginelli et al., 2007) diagnose asymptotic growth; Jacobian regularization tunes propagation directly (Engelken, 2023). The present block decomposition instead identifies where a measured defect acts. Classical invariant-subspace perturbation theory controls subspace rotation using spectral separation or residual size (Davis & Kahan, 1970; Stewart, 1973). Our result is complementary: it resolves four tangent/normal routes for a generally nonnormal recurrent map and converts them into memory-rate bounds.

Noise projection near cycles and tori (MacLaurin, 2023; Ahsan et al., 2025) and drift–diffusion matching (Nartallo-Kaluarachchi et al., 2026) are complementary forward problems. We ask the inverse local design question: for a fixed action, decoder, and ambient noise, which decoded covariance is unavoidable? The answer is conditional on those choices and does not assert a universal representation optimum.

Input-driven continuous-attractor RNNs also use invariant manifolds for representation transfer and domain adaptation (Xu et al., 2026). Their goal differs from ours: we ask when exact per-input equivariance transports recurrent memory tangents and how measured defects and state noise act on the decoded coordinates. Temporal-scaling equivariant circuits encode a group parameter through a control input to generate sequences at different speeds (Zuo et al., 2023); our controlled-cocycle result instead concerns pathwise tangent transport and finite-horizon degradation.

Formal verification of neural networks is developing rapidly. TorchLean gives neural-network definitions, execution, finite-precision semantics, and verification a shared Lean semantics (George et al., 2026). Our development instead machine-checks selected dynamical bounds and architecture-level equivariance, while correspondence to the released PyTorch checkpoints is tested executably rather than formalized as PyTorch semantics. Formalized Hopfield and Boltzmann models provide a complementary Lean treatment of classical recurrent-memory convergence and stochastic ergodicity (Cipollina et al., 2025).

M LEARNED-TASK AND BASELINE SCOPE CHECKS

The earlier S^1 learned experiment grounds the framework in a recognizable continuous-memory task while remaining application evidence rather than theorem proof. The exact cell has state (z, h) with $z \in \mathbb{R}^2$ transforming by rotation and h invariant. Its vector field has form

$$\dot{z} = a(\|z\|^2, h, u)z + b(\|z\|^2, h, u)Jz, \qquad \dot{h} = g(\|z\|^2, h, u),$$

which is equivariant for each scalar invariant control u. The zero-input restriction is covered subject to the autonomous theorem's hypotheses, and the per-input theorem in Theorem 1 now also gives exact tangent transport for arbitrary input sequences as long as the orbit rank remains fixed. Task error is still consequence evidence, because good decoding and transverse stability are additional requirements.

The matched scope checks use GRU (Cho et al., 2014), LSTM (Hochreiter & Schmidhuber, 1997), and orthogonal or unitary recurrent baselines (Henaff et al., 2016; Arjovsky et al., 2016). They provide task-level context under the reported budgets rather than a claim that generic gated or norm-preserving recurrent networks cannot integrate.

The archived comparison uses training horizon 64, batch size 64, AdamW and gradient clipping at 1. The generic-baseline reruns use learning rate 0.002; the inherited equivariant reference uses 0.003. Thus these are controlled budget comparisons, not identical-hyperparameter retraining of all models. The per-seed records, configurations and table-generation command are included under `empirical_provenance/baseline_scope/`.

Table 14: Original and stronger generic-baseline check on the horizon-256, speed-1.8 full-phase slice. The stronger baselines use hidden size 32 and 500 optimization steps. Entries are mean ± SEM over three seeds per baseline configuration and six seeds for the equivariant reference.

Configuration	Full-phase circular RMSE	Restricted-phase RMSE
Equivariant reference	0.2554 ± 0.016	–
GRU, original budget	1.748 ± 0.023	1.778 ± 0.008
GRU, stronger budget	1.212 ± 0.026	1.619 ± 0.019
LSTM, original budget	1.730 ± 0.021	1.801 ± 0.020
LSTM, stronger budget	1.156 ± 0.029	1.678 ± 0.032
Orthogonal RNN, original budget	1.790 ± 0.008	1.805 ± 0.012
Orthogonal RNN, stronger budget	1.749 ± 0.006	1.789 ± 0.012

GRU and LSTM validation curves were still improving at the end of the original 120-step budget and improve substantially under the stronger budget. These comparisons support an inductive-bias and sample-efficiency advantage under the tested settings, not an inability of generic gated RNNs to integrate. Extended sweeps and diagnostic plots are supplied as task-level context; they are not assumptions or proofs of the theoretical results.

N REPRODUCIBILITY

The anonymous supplement provides the Python and Lean sources, pinned dependencies, study configurations, checkpoint registries, recorded random-number schedules, result tables, and numerical evaluation records. Figure sources include generation scripts and the documented label-only correction for the reduced phase-field plot. The supplement's `README_REPRODUCIBILITY.md` maps the scientific studies to their code, data, and reproduction commands. Historical six-unit records remain separate from the independent replication, as described in Section J.2.

The executable checks cover equivariance and tangent transport, complete training-unit and condition accounting, local and finite-horizon bounds, decoder covariance, and paired task summaries. Unsuccessful preliminary runs and out-of-regime cases are retained. Random-stream separation is documented in Section K, and the higher-dimensional study's operator-norm correction is explained in Section K.4. Source hashes, dependency versions, complete commands, and dated build records are supplied with the corresponding code rather than treated as additional empirical replicates.

N.1 FORMAL VERIFICATION AND SCOPE

The supplementary Lean 4 project includes pinned Lean and Mathlib dependencies, a manuscript-to-declaration mapping, and recorded build and `-trust=0` declaration checks. The dependency report lists only Lean/Mathlib foundational axioms; the source scan finds no project `sorry`, `admit`, or declared `axiom`. Table 15 distinguishes the encoded results from mathematical interfaces and numerical implementation checks.

Table 15: Formal verification scope. "External" means used with explicit assumptions or a cited theorem, not claimed as a Lean result.

Manuscript component	Formal status	Notes
Specified finite graph transport	Lean proved	Unique forward sequence for the supplied initial graph under the denominator condition.
Finite center-product bounds	Lean proved	Pointwise lower and upper products in chronological order.
Endpoint graph factors	Lean proved	Includes the Euclidean $\sqrt{1+r^2}$ factor and frame endpoints.
Subspace-angle control	Lean + manuscript	Lean proves directed-gap sine/arcsine bounds; Lemma 11 identifies the conventional largest principal angle.
Leakage scaling	Lean proved	Scalar recurrence gives the stated $O(\epsilon^2/d)$ bound under its smallness assumptions.
Symbolic S^1 cell equivariance	Lean proved	Invariant coefficients imply exact field and Euler-step equivariance.
Group-tangent transport	Lean proved	Differentiation and cocycle telescoping are machine checked.
Direct continuous-time neutrality	Lean proved	For a finite-dimensional local action parameterization with uniform tangent bounds.
General Lie-orbit construction	External	Smooth action and Lie-exponential orbit curves use standard equivariant dynamics.
Full-spectrum multiplicity	External	The at-least-q corollary invokes an applicable multiplicative ergodic theorem.
PyTorch/checkpoint correspondence	Executable	Strict loading and numerical residual checks; not a Lean theorem about PyTorch semantics.
Stochastic decoder covariance	Manuscript proof	The linear-algebra result is proved in the appendix but not formalized in Lean.
Empirical coverage and task claims	Executable	Recomputed from released row-level tables; not theorem evidence.

For the learned S^1 family, the formal development connects symbolic equivariance to tangent transport and zero direct-tangent growth under noncollapse and uniform action bounds. The continuous-time result uses a finite-dimensional local action parameterization, without a multiplicative ergodic theorem. The broader full-spectrum conclusion uses the external theorem under the assumptions stated in Corollary 2.

Separate Python checks compare the declared S^1/T^2 formulas with stored checkpoints through strict loading, parameter counts, sampled equivariance, and analytical tangent transport. The largest discrepancy in the 24 recorded moving-frame rows was 1.11×10^{-15}, within floating-point roundoff. These checks do not formalize PyTorch execution semantics. The stochastic decoder analysis, including anisotropic radius allocation, is proved in the manuscript; the bounded-radial T^4/T^8 implementation is tested numerically and is not covered by the Lean proof.

N.2 COMPUTATIONAL RESOURCES

The independent replication and higher-dimensional study used a CPU environment with Python 3.12.14, NumPy 2.4.6, and PyTorch 2.5.1+cpu on an AMD Ryzen 9 9950X with 16 cores and 125.6 GiB of installed memory. The recorded sum of run-wall times was 717.63 s for the 100 T^2 replication runs and one disjoint preliminary run, 55.60 s for the four initial higher-dimensional preliminary runs, and 4,511.03 s for the 192 higher-dimensional evaluation runs. The evaluation used one PyTorch CPU thread. These are sums for the stated run groups, not the total computational cost of the project; repeated evaluation and additional checks are recorded separately. Peak memory was not instrumented. Full software pins, runtime records, and reproduction commands are included in the supplement.