Supplementary Material for: State Transitions Unlock Temporal Memory in Swarm-Based Reservoir Computing

Tanner Lund, Alyssa Adams, Nathanael Aubert-Kato, and Takashi Ikegami

Terminology note. The main text uses *single-state* (no transitions; pure-MC protocol) and *two-state* (with transitions). Some legacy figures below retain historical labels *Naive*, *Integrated*, and *Multi-state*. Here, *Integrated* refers to a tuned *single-state* pipeline from earlier drafts that used polynomial features; it is *not* identical to the *single-state (pure MC)* condition used in our ablations (which uses linear readout without polynomial terms). We include these legacy visuals for completeness; current results and claims are based on the single-state vs two-state toggle under the pure-MC protocol.

S1 Supplementary Methods

S1.1 Performance Metrics

We report standard forecasting metrics consistent with RC practices [1, 2, 3]:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2, \tag{1}$$

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\max_i\, y_i - \min_i\, y_i}, \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}. \tag{3}$$

We reproduce these standard definitions here for completeness, as they are used in both the Mackey-Glass forecasting and memory-capacity summaries. MSE measures the average squared prediction error in the original units of the signal, NRMSE rescales this error relative to the signal range (so that different tasks are comparable) , and R^2 measures the fraction of variance in the target explained by the model (with $R^2 = 1$ indicating perfect prediction and $R^2 = 0$ matching a constant-mean baseline) .

S1.2 Echo State Property (ESP) Validation

The echo state property ensures that reservoir states asymptotically depend only on recent driving inputs rather than initial conditions [4]. We validate ESP through systematic perturbation analysis.

S1.2.1 Perturbation Protocol

- **Warmup period**: 50 timesteps to establish baseline dynamics
- **Perturbation window**: 20 timesteps (steps 50-70)
- **Recovery period**: 50 additional timesteps
- **Perturbation intensities**: -1.0, -0.5, 0.0, 0.05 (applied as temperature multipliers)

S1.2.2 Metrics Analyzed

- **Mean Position Distance**: Captures spatial configuration stability
- **Local Order**: Measures the average local alignment in the swarm.
- **Velocity Variance**: Reflects kinetic energy distribution changes

A system demonstrates ESP if these metrics converge after perturbation ends, indicating that transient effects decay and the system returns to input-driven dynamics.

S1.3 Feature Set and Readout

Permutation-invariant statistics. The reservoir state is computed from permutation/rotation/translation-invariant swarm statistics, including: convex-hull area (CH), mean pairwise distance (APD), mean unit-velocity alignment (AA), mean speed (AS), speed standard deviation (σ_v), angular momentum about the centroid (AM), fraction of agents in clustered state (CR), and mean speeds by state (AS_0, AS_1). Features are standardized (zero mean, unit variance) before readout training; PCA-whitening is available in the codebase but was not used in the canonical runs reported here.

Hull-area operator (convex hull area) For completeness, we give the explicit form of the convex-hull area operator used for $f_{\mathrm{CH}}(t)$ in the main text. Let $\mathrm{Hull}(\{\mathbf{x}_i(t)\})$ have M vertices in counterclockwise order, $\mathbf{v}_1(t), \ldots, \mathbf{v}_M(t)$, with $\mathbf{v}_m(t) = (x_m(t), y_m(t))$ and $\mathbf{v}_{M+1}(t) = \mathbf{v}_1(t)$. Then

$$\mathrm{AreaHull}(t) = \begin{cases} \dfrac{1}{2}\left|\displaystyle\sum_{m=1}^{M}\big(x_m(t)y_{m+1}(t) - x_{m+1}(t)y_m(t)\big)\right|, & M \geq 3, \\ 0, & M < 3. \end{cases}$$

Pure-MC readout (used for ablation and slope confirmation). Ridge regression with cross-validated regularization (RidgeCV over $\alpha \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}\}$ with R^2 scoring on the training segment); no polynomial terms; standardized features. Regression targets are standardized within the training pipeline for numerical stability across delays.

S1.4 ESN Configuration and Protocol (Baseline)

We include a canonical ESN as contextual baseline. Reservoir sizes: small= 600, medium= 2400, large= 4800 units. Hyperparameters: spectral radius 0.95; sparsity $\in [0.90, 0.98]$ by size; input scaling 0.3; leak rate 0.3. Readout: ridge regression with cross-validated regularization (RidgeCV over $\alpha \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}\}$ with R^2 scoring on the training segment). Regression targets are standardized within the training pipeline for numerical stability. MC protocol matches pure-MC (i.i.d. input; washout 500; delays 1–50; linear readout). Means and 95% CIs over n=5 seeds per size are reported in Supplementary Results.

ESN scaling (summary). Total MC increases with reservoir size; a simple least-squares fit on per-size means yields an approximately linear trend (slope positive; intercept near zero), consistent with classical RC expectations. Exact per-size means and 95% CIs are listed in Supplementary Results. As a compact summary (means $\pm$ 95% CI, n=5): small (600): 3.71 ± 0.04, medium (2400): 11.65 ± 0.72, large (4800): 16.26 ± 0.71.

S1.5 Closed-Loop Evaluation

We perform a compact autonomous (closed-loop) check in which model outputs are fed back as inputs after a warmup. This tests whether the reservoir has internalized target dynamics beyond short-term mapping.

Following established reservoir computing protocols [5, 2], we evaluate each model's ability to autonomously generate Mackey-Glass dynamics when its own predictions are fed back as inputs.

S1.5.1 Closed-Loop Prediction Protocol

We evaluate each model's ability to generate stable autonomous predictions using a continuous prediction protocol. After training on the Mackey-Glass time series, we test how long each model can maintain accurate predictions when operating in a fully autonomous mode. The protocol consists of:

- **Warmup Phase (200 steps):** The model receives ground truth inputs to initialize the swarm state.
- **Autonomous Prediction Phase (up to 2000 steps):** The model recursively uses its own predictions as inputs, with no ground truth correction.

During the prediction phase, we compute the Normalized Root Mean Square Error (NRMSE) at each timestep by comparing the model's predictions to the ground truth continuation of the Mackey-Glass series. We use the global training data range for normalization to ensure consistent scaling. A model is considered to have diverged when NRMSE exceeds 0.3, indicating predictions have deviated substantially from the true dynamics. This protocol provides a test of dynamical stability, as small prediction errors can compound over time, leading to divergence from the true attractor.

S2 Supplementary Results

S2.1 ESP Validation Results

Representative perturbation tests are shown here for brevity. Full code and configurations are available with the repository; quantitative summaries indicate convergence after perturbations consistent with ESP. Representative perturbation responses for legacy pipelines are shown in Supplementary Figures S1–S3.

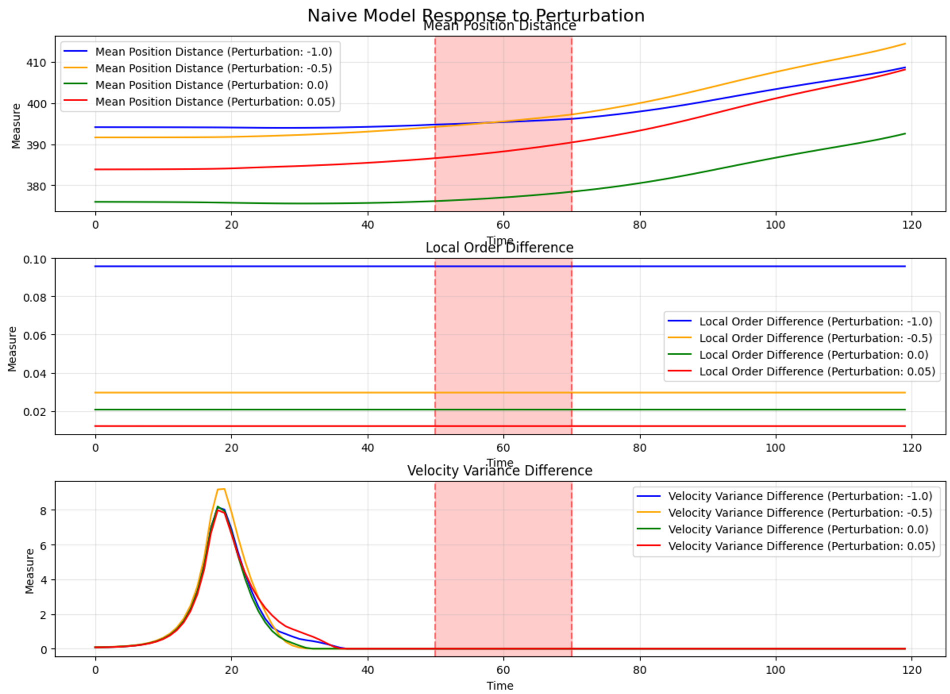

Figure S1: **Legacy.**
Perturbation test results for a legacy naive pipeline. Red shaded area indicates perturbation period (timesteps 50–70).
Terminology mapping: "Naive" denotes a historical baseline; current terminology elsewhere uses *single-state* and *two-state*.

S2.2 Closed-Loop Prediction Results

Table S1: **Closed-loop NRMSE performance metrics**

Model	Mean Steps to Divergence	Std Dev	Mean Final NRMSE
Naive	2000*	0	0.190 ± 0.008
Integrated	1080	921	0.268 ± 0.037
Multi-State	2000*	0	0.235 ± 0.017

*No divergence (NRMSE $>$ 0.3) observed within the 2000-step test period; values averaged over 10 trials

Legacy closed-loop trajectories and NRMSE traces are shown in Supplementary Figures S4–S6, with summary statistics in Supplementary Table S1.

Note on interpretation. These autonomous (closed-loop) checks probe qualitative stability once predictions are fed back as inputs. They are not directly comparable to open-loop Mackey – Glass forecasting scores reported in the main text, which assess one-step accuracy with ground-truth inputs. We include closed-loop plots to illustrate dynamical behavior only.

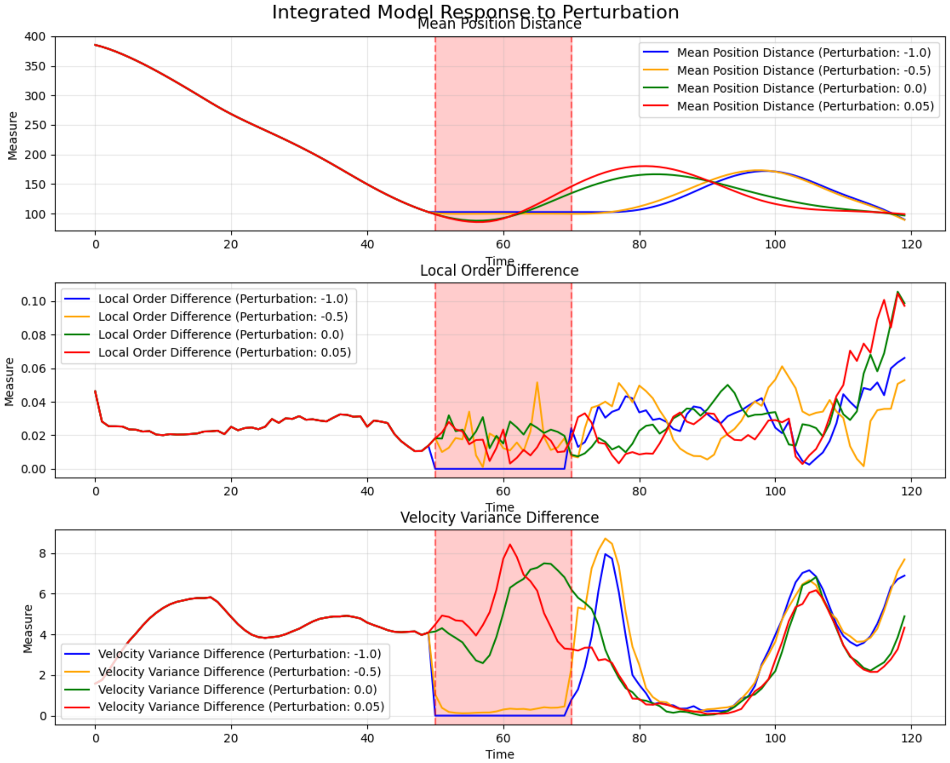

Figure S2: **Legacy.**
Perturbation test results for a legacy integrated pipeline (single-state with polynomial features). Terminology mapping: "Integrated" corresponds to *single-state*.

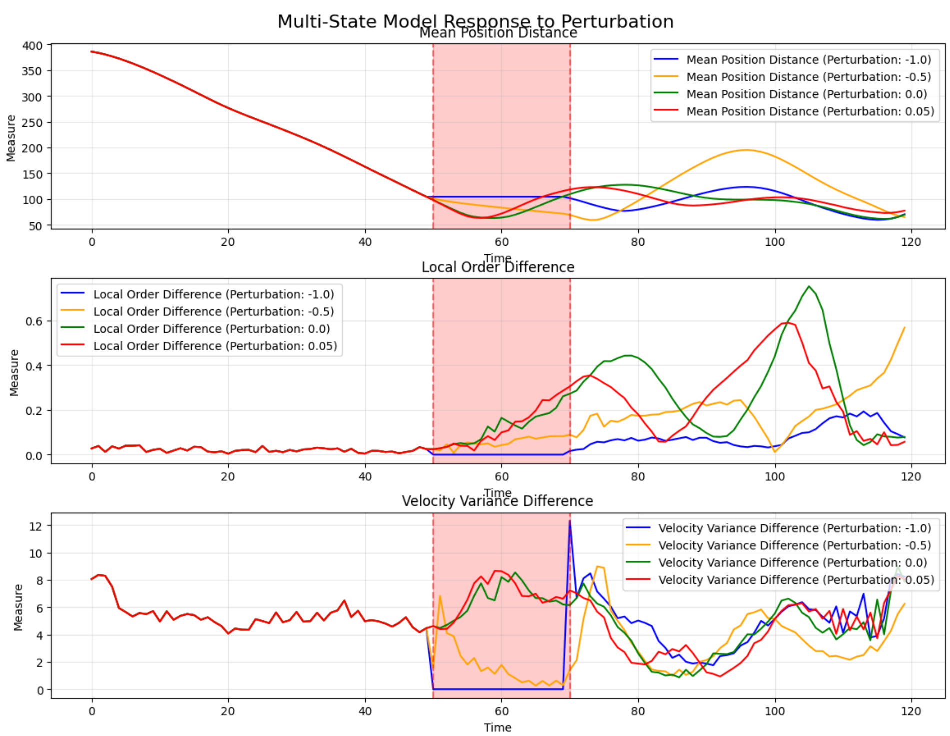

Figure S3: **Legacy.**
Perturbation test results for a legacy multi-state pipeline. Terminology mapping: "Multi-state" corresponds to *two-state*.

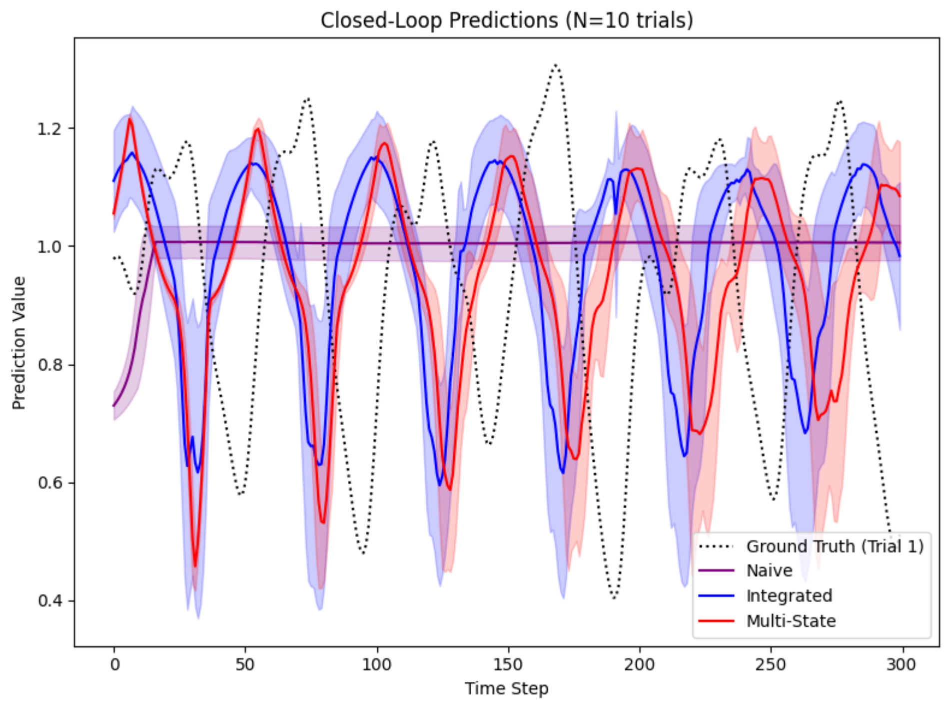

Figure S4: **Legacy.**
Closed-loop predictions for legacy pipelines. Terminology mapping: "Integrated" → *single-state*; "Multi-state" → *two-state*.

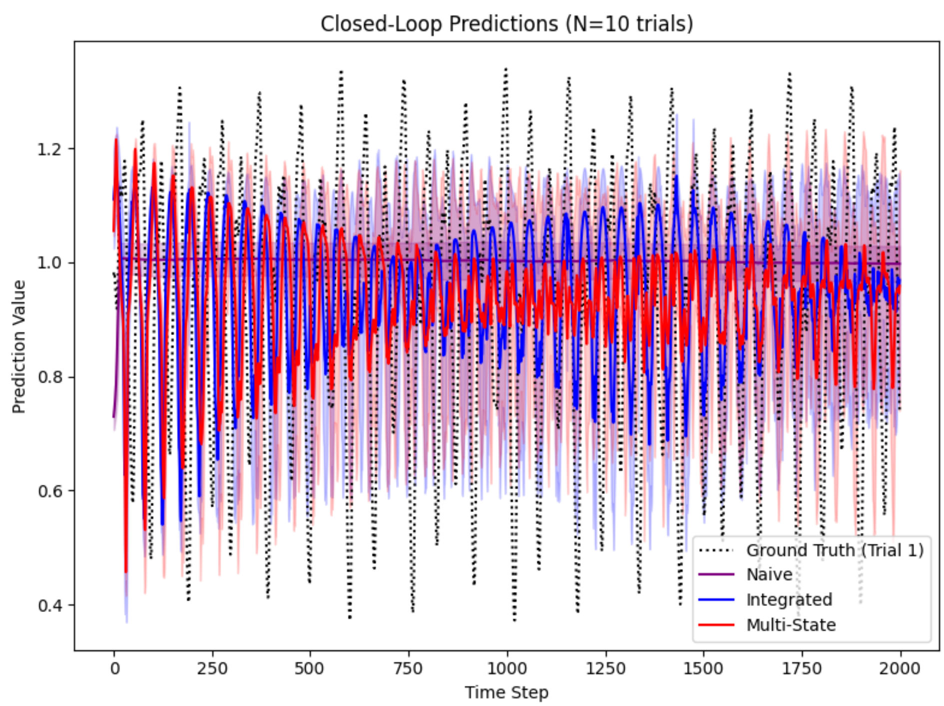

Figure S5: **Legacy.**
Closed-loop predictions for 2000 steps. Terminology mapping: "Integrated" → *single-state*; "Multi-state" → *two-state*.

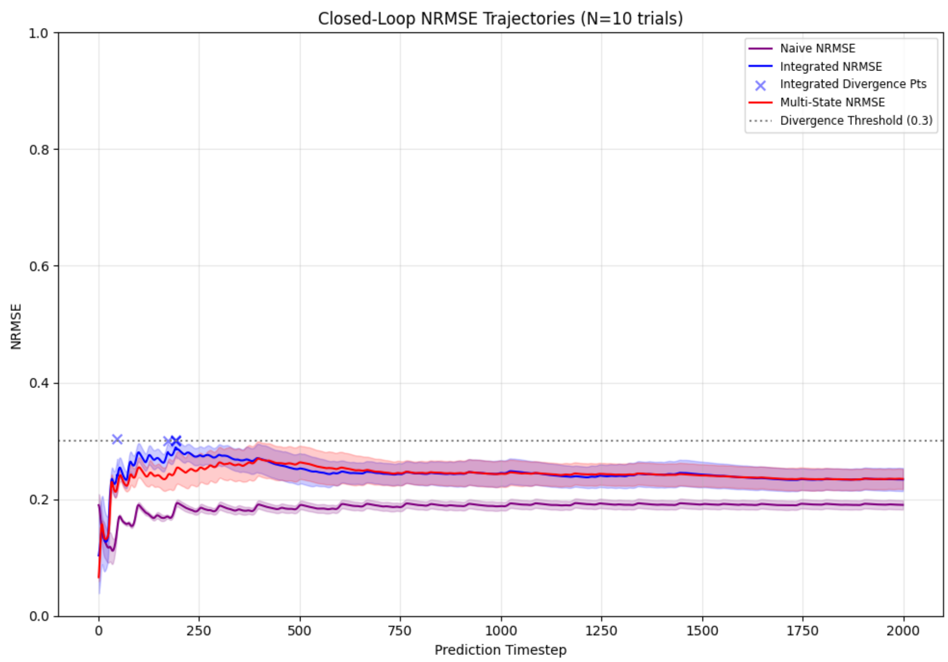

Figure S6: **Legacy.**
Closed-loop NRMSE scores. Terminology mapping: "Integrated" → *single-state*; "Multi-state" → *two-state*.

S2.3 Mackey-Glass (Open-Loop) Verification

Open-loop one-step forecasting on Mackey – Glass (delay τ=17) confirms strong single-state performance (mean ± std over n=10 seeds):

- Single-state: $R^2 = 0.9609 \pm 0.0018$, NRMSE $= 0.1976 \pm 0.0045$ (n=10).
- Two-state: $R^2 = -2.76 \pm 4.50$, NRMSE $= 1.675 \pm 0.978$ (n=10).

Under the same protocol, the two-state configuration performs poorly, consistent with the main text's focus on memory rather than MG forecasting for two-state swarms.

S2.4 ESN MC Baseline (Context)

For context, a canonical ESN exhibits the expected increase of total MC with reservoir size. Mean ± 95% CI over 5 seeds:

- small (600): 3.71 ± 0.04
- medium (2400): 11.65 ± 0.72
- large (4800): 16.26 ± 0.71

These values provide a neural baseline for magnitude and scaling; see the main text for discussion.

S2.5 Pure MC Scaling (Finalized Subset and Extended Coverage)

Across merged small+large-N data, the linear-readout, no-polynomial "pure MC" protocol yields a slope in the ∼0.012–0.014 per-boid band over N=2–2000 (fit on per-N means), consistent with the main-pipeline scaling. The finalized subset results at two sizes (used in the main text previously) are:

- N=800: total MC $= 11.42 \pm 0.08$ (95% CI), n=10
- N=1600: total MC $= 21.21 \pm 0.03$ (95% CI), n=10

Fit on these per-N means: MC $\approx 0.0122 \cdot N + 1.64$.

Extended coverage. Merging the small-N set with additional large-N pure-MC runs (including N=1200, 2000; n=10 each, σ=0) yields a consistent slope (Figure S7). The main text displays the merged large-N pure-MC panel at N=800, 1200, 1600, 2000 alongside the main-pipeline scaling; here we retain the small-N evidence and the merged small+large-N panel for completeness.

S2.6 Density and small-N behavior

At very small populations, neighbor interactions are sparse and the local-energy trigger rarely crosses thresholds in an informative way, reducing linearly decodable memory. A small-N sweep under the pure-MC protocol shows the expected bend below N≈20–30 (Figure S8), while merged small+large-N data recover the large-N slope. A density-controlled check that holds the expected neighbor count approximately constant across N lifts the small-N points and restores linearity, indicating the bend is a sparsity effect rather than a failure of the multi-state mechanism (Figures S9-S10).

S2.7 Short-Delay $\mathbf{R}^2$ (Pure MC)

For an exemplar one-seed short-delay table per N and the corresponding mean ± 95% CI summaries, see Supplementary Tables S2 and S3, respectively.

Table S2: **Exemplar one-seed short-delay $\mathbf{R}^2$ per N (pure MC).**

N	R_1^2	R_2^2	R_3^2	R_4^2
800	0.898	0.383	0.304	0.250
1200	0.921	0.464	0.410	0.384
1600	0.933	0.582	0.513	0.473
2000	0.924	0.652	0.601	0.599

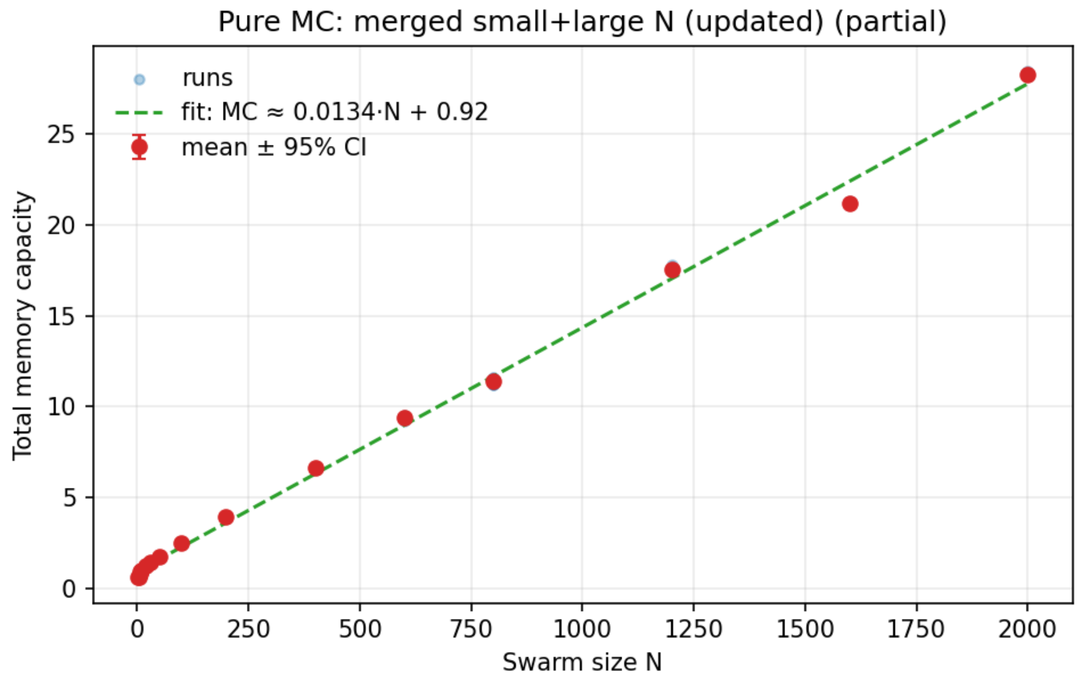

Figure S7: Pure-MC scaling
(merged small + large N) . Points show per-seed totals; lines show fit on per-N means.

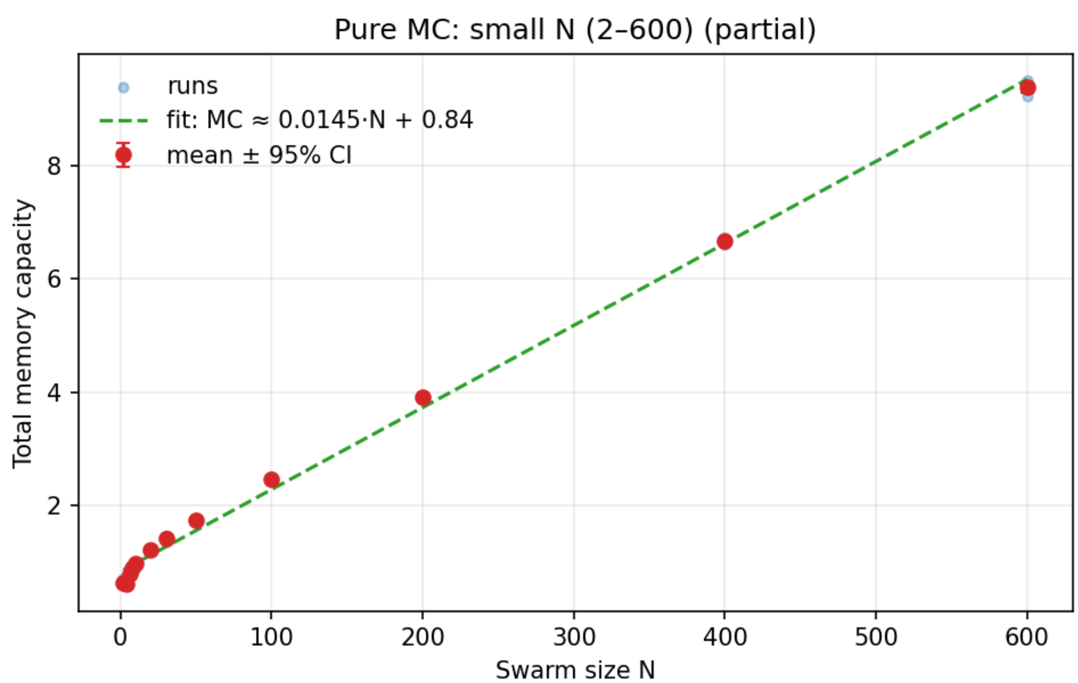

Figure S8: Pure-MC scaling at small N. Linear law holds from ∼20–30 upward; below that range the curve bends due to sparse interactions.

Table S3: **Short-delay $\mathbf{R}^2$ (mean ± 95% CI, n=10 seeds).**

N	R^2_1	R^2_2	R^2_3	R^2_4
1200	0.936 ± 0.007	0.499 ± 0.021	0.420 ± 0.011	0.373 ± 0.007
1600	0.932 ± 0.005	0.561 ± 0.015	0.496 ± 0.010	0.458 ± 0.006

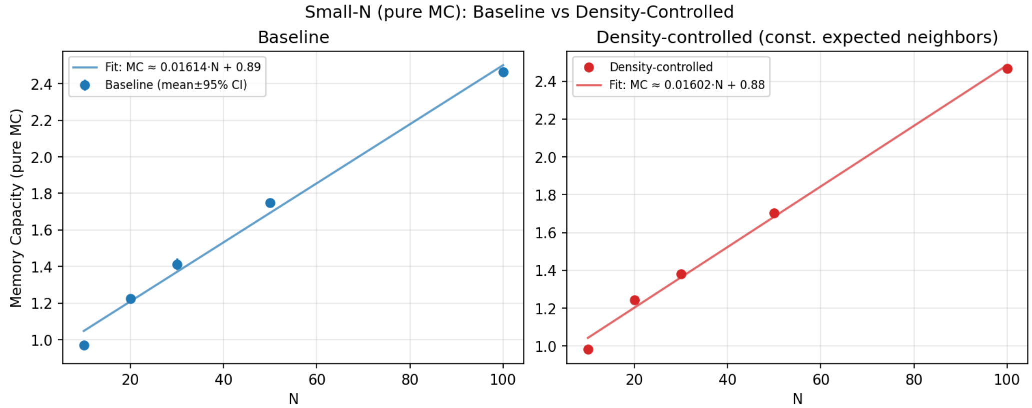

Figure S9: Pure-MC scaling at small N
(baseline vs density-controlled) . Holding neighbor density approximately constant lifts small-N points and restores linearity.

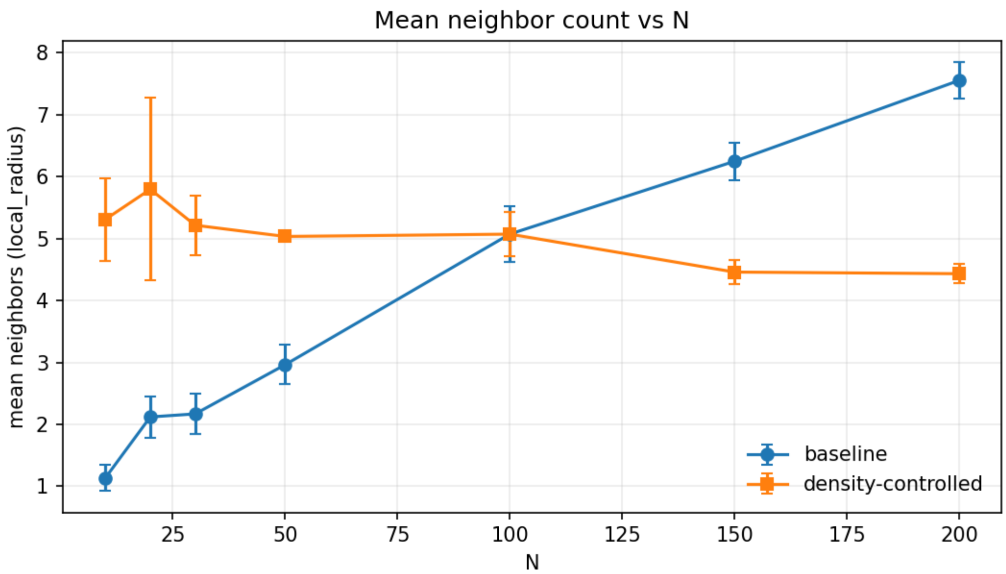

Figure S10: Expected neighbors vs N under baseline and density-controlled settings. Density control holds neighbor counts roughly constant across N.

S2.8 Ablation at Fixed N (Pure MC)

Under the pure MC protocol, two-state swarms vastly exceed single-state in total MC:

- N=800: 0.11 ± 0.02 vs 11.43 ± 0.11 (n=10 each)
- N=1200: 0.07 ± 0.01 vs 16.33 ± 0.07 (n=10 each)
- N=1600: 0.11 ± 0.01 vs 21.19 ± 0.05 (n=10 each)

S2.9 Long-trajectory reliability at N=800

We report a compact reliability check under long trajectories using a sliding window MC and state-activity traces. Two operating points are contrasted: (i) baseline defaults that can drift to an all-dispersed absorbing state, and (ii) a *brief-burst* corridor tuned via thresholds (e.g., ρ=1.70, h=0.10) that maintains low, nonzero clustered occupancy over time.

- Baseline defaults (seeded runs at $\sigma \in \{0, 0.1\}$) : the collapse proxy flags all runs as collapsed to all-dispersed (state 0) with first flagged window near index ∼2000; activity traces show near-zero clustered fraction thereafter.
- Tuned brief-burst corridor (ρ=1.70, h=0.10; $\sigma \in \{0, 0.1\}$) : no collapses observed across seeds under our proxy; mean clustered fraction remains low but nonzero (brief excursions) ; windowed MC remains stable.

Those results are summarized in Table S4.
See Supplementary Figures S11–S12 for representative windowed-MC traces under baseline and tuned settings.

Table S4: **Aggregate reliability metrics (mean across seeds) for $\sigma \in \{0, 0.1\}$**

Setting	Collapse frequency	Time to first collapse	Mean frac(state 1)
Baseline defaults (σ=0)	100%	∼2000 (index midpoint)	≈ 0.00
Baseline defaults (σ=0.1)	100%	∼2000 (index midpoint)	≈ 0.00
Tuned (ρ=1.70, h=0.10, σ=0)	0%	N/A	≈ 0.12
Tuned (ρ=1.70, h=0.10, σ=0.1)	0%	N/A	≈ 0.12

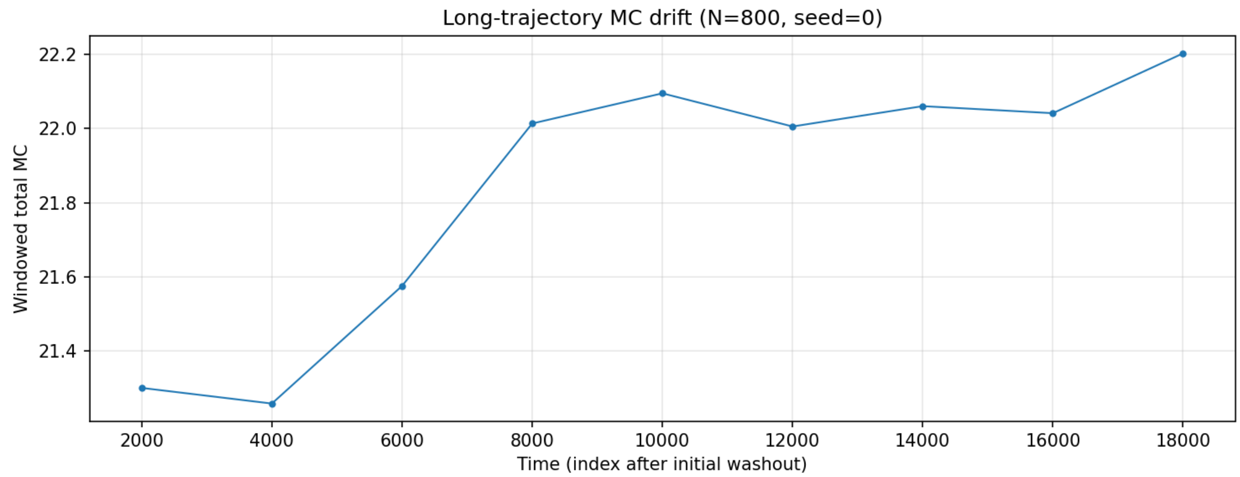

Figure S11: Windowed total MC
(4k window; 2k stride) at N=800 under baseline defaults. In this long-trajectory path (enriched multi-state features) , MC remains stable even as activity regimes differ; canonical pure-MC slope metrics use invariant features (see main text) .

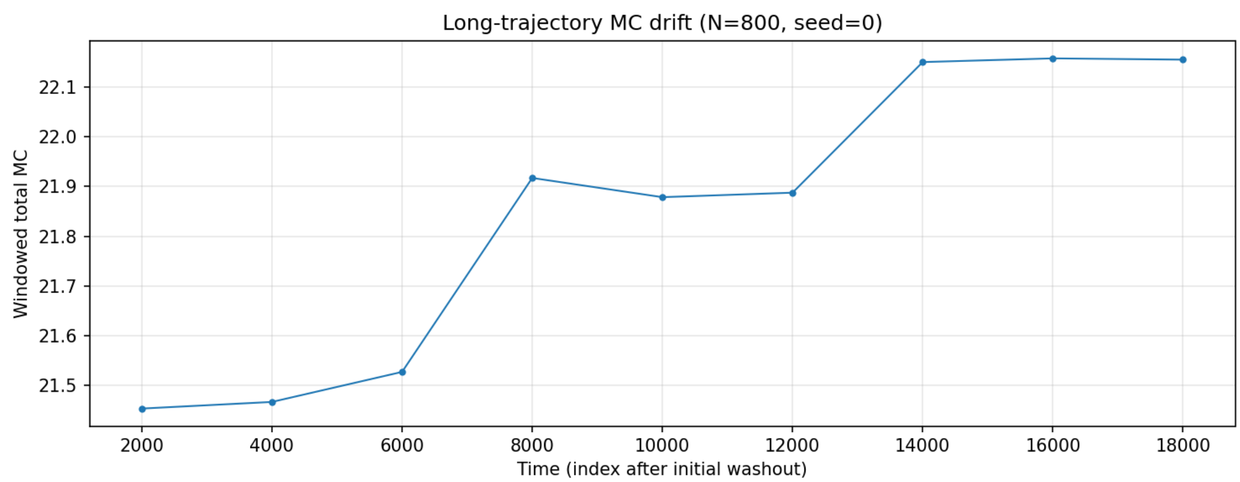

Figure S12: Windowed total MC
(4k window; 2k stride) at N=800 under tuned brief-burst thresholds.

S2.10 Main Scaling Summary (Two-State)

Across N=800–2000 (five seeds per noise, three noises; n=15 per N), total MC scales linearly with N:

- Per-N means ± 95% CI: 800: 11.41 ± 0.08; 1200: 16.36 ± 0.05; 1600: 21.22 ± 0.05; 2000: 26.14 ± 0.06
- Fit on per-N means: MC $\approx 0.0123 \cdot N + 1.61$; slope 95% CI: [0.0122, 0.0123]

S2.11 Parameter sensitivity at N=1200

We tested ±15% one-factor perturbations around a tuned operating point (activity threshold ρ=1.70, hysteresis factor h=0.10) for two representative dynamics knobs: dispersed outer radius and dispersed separation gain (FigureS13). Under the long-trajectory feature path, total MC remained stable across scales with varying activity fractions and no abrupt discontinuities, indicating a flat sensitivity corridor in this regime. This robustness supports the long-trajectory reliability interpretation in Section S2.9, where brief-burst settings maintain low, nonzero clustered occupancy without collapse across seeds.

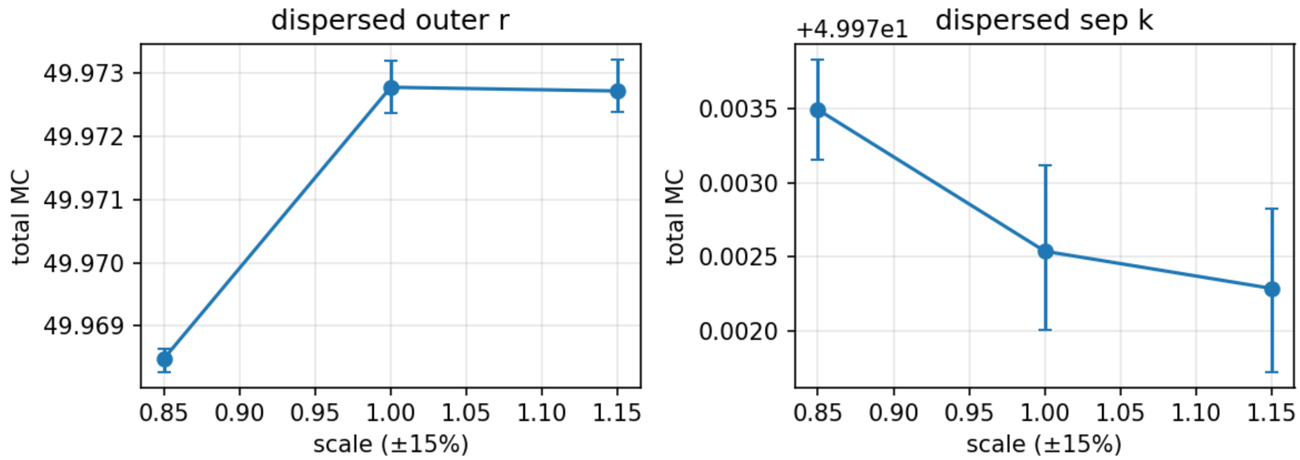

Figure S13: One-factor ±15% sensitivity at N=1200 for dispersed outer radius (left) and dispersed separation gain (right) ; error bars show 95% bootstrap CIs on the mean. Total MC remains stable across scales in this regime. (MC is near the ceiling $\tau_{\max}$=50 in this long-trajectory path; variations are on the order of 10^{-3}–10^{-2}.)

Note: under this long-trajectory feature path the measured total MC is near the protocol ceiling $\tau_{\max}$=50, so the y-axis variation is extremely small (sub-percent; see plotted scale). These plots demonstrate robustness (flat sensitivity corridors) rather than meaningful differences between knob settings.

S2.12 Parameter sensitivity at N=1600

Mirroring the N=1200 analysis, ±15% one-factor perturbations around the tuned operating point at N=1600 showed stable total MC across scales for dispersed outer radius and dispersed separation, with no abrupt discontinuities (Figure S14). These results further align with the reliability trends in Section S2.9. To explore further the small range of changes in total MC, we provide a plot of ΔMC changes compared to baseline parameters (Figure SS15). We also report a narrow α-multiplier check around the nominal operating point (Figure SS16); while modest changes in α near the brief-burst corridor have negligible impact at this resolution, a broader sweep reveals a non-monotone dependence with degraded MC at extreme α (see text below).

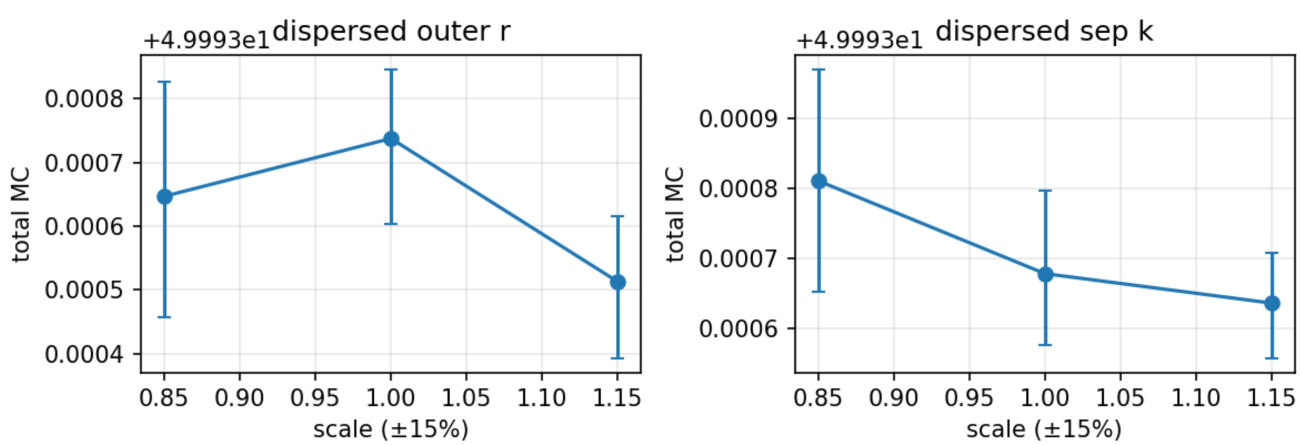

Figure S14: One-factor ±15% sensitivity at N=1600 for dispersed outer radius (left) and dispersed separation gain (right) ; error bars show 95% bootstrap CIs on the mean. (MC is near the ceiling $\tau_{\max}$=50 in this long-trajectory path; variations are on the order of 10^{-3}–10^{-2}.)

As with N=1200, MC is near the $\tau_{\max}$=50 ceiling here; observed differences are sub-percent and these plots serve to illustrate robustness rather than effect sizes.

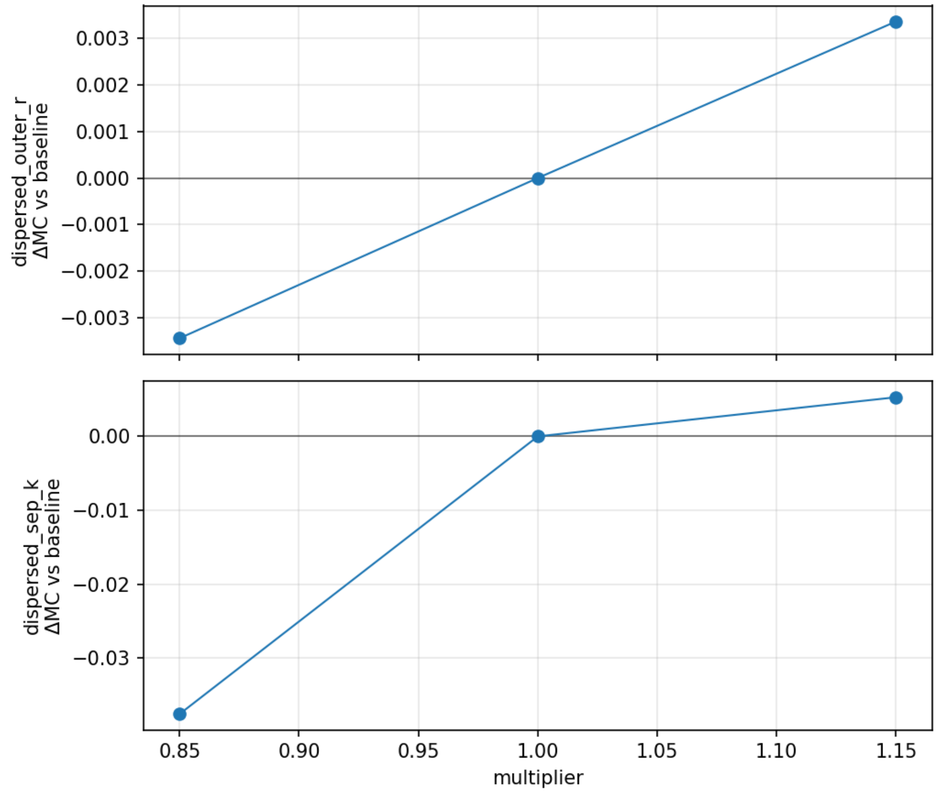

Figure S15: Complementary pure-MC check at N=1600 (linear readout; invariant features; steps=4000; seeds=3): ΔMC relative to the nominal setting for $\pm15\%$ one-factor changes in dispersed outer radius and dispersed separation. Deltas are small compared to short-run variability, consistent with a flat sensitivity corridor at large N.

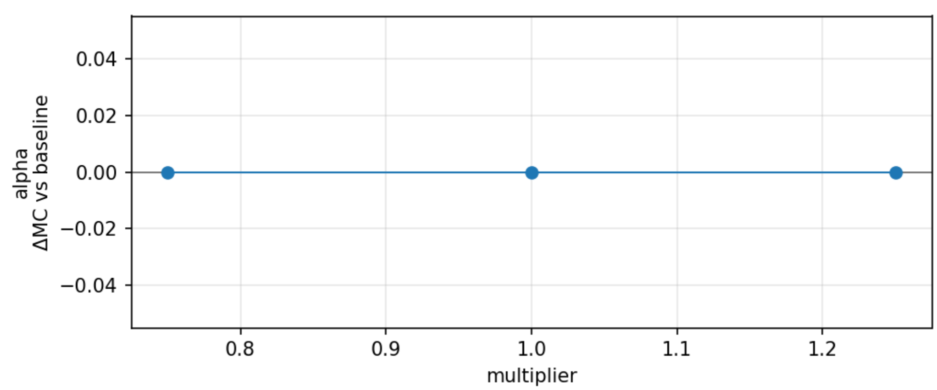

Figure S16: ΔMC at N=1600 under the pure-MC protocol for α multipliers $\{0.75, 1.0, 1.25\}$ (steps=4000; seeds=3). Deltas are effectively zero at this resolution, supporting that modest changes in α within the brief-burst corridor do not materially affect total MC at large N, though a broader sweep reveals a non-monotone dependence with degraded MC at extreme α.

Extended α sweep. To directly address whether α can qualitatively change memory, we also ran a broader pure-MC sweep at N=800 under the same delay window ($\tau_{\max}$=50) with steps=8000, washout=500, energy threshold ρ=1.70, hysteresis factor h=0.10, input mode 'temperature', invariant-9 features, and seeds $\{0, 1, 2\}$. Total MC varies strongly and non-monotonically with α: the peak mean occurs near $\alpha\approx0.3$ (MC = 0.527 ± 0.058, 95% CI; n=3), while a minimum occurs near $\alpha\approx0.8$ (MC = 0.092 ± 0.028), a best/worst ratio of $\sim 5.75\times$. Relative to the manuscript nominal α=0.4, α=0.3 is +0.059 absolute MC (about +12.6%) under this pilot configuration. These results clarify that while small perturbations around the nominal corridor may have negligible effect, α is not purely a time-rescaling parameter over a broad range and extreme values can move the system away from a high-performance regime.

S2.13 Non-optimal regimes

(illustrative) To contextualize the brief-burst operating corridor, we include examples of regimes that degrade MC:

- **High clustered occupancy** (persistent red): MC decreases when state s_1 occupancy is high/persistent relative to brief, input-aligned excursions. See the trade-off chart relating MC to activity measures (Fig. S17) .
- **Near-symmetric state fractions**: settings that hover near balanced state fractions also underperform relative to brief-burst regimes.

These patterns are consistent with the mechanism that short-lag, slow variables are induced by intermittent transitions (Figure S18) rather than persistent clustering (Figure S19).

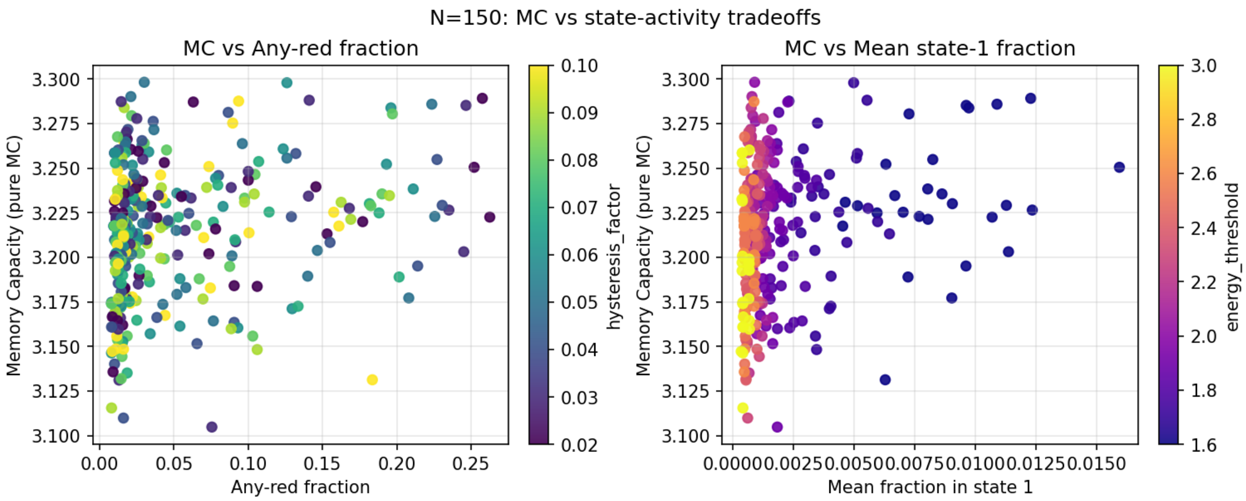

Figure S17: Trade-off at N=150: MC vs activity statistics across threshold/hysteresis grids. Near-optimal MC occurs with low mean state s_1 and intermittent bursts; high/persistent state s_1 reduces MC.

S2.14 Slow-variables quantification

(representative) In addition to the representative time-series panel (Figure S20), we examined cross-correlations between $u(t-\tau)$ and the slow, state-aware features. Observed correlations are small and not significant under a max-over-lags circular-shift baseline, but their strongest structure is concentrated at small delays (order of a few steps), consistent with short-lag responses without requiring persistent clustered occupancy. This aligns with the qualitative “brief-burst” picture and avoids a causal interpretation.

Notably, performant regimes do not require large, persistent encapsulation structures; intermittent, input-aligned excursions into state s_1 suffice to generate the measured slow-variable responses.

Because $u(t)$ is i.i.d. under brief-burst driving, alignment is assessed via cross-correlation rather than visual co-fluctuation. Figure S21 shows cross-correlations between $u(t-\tau)$ and the three features over small lags; the strongest structure is concentrated at small delays, consistent with the short-lag response claim.

S2.15 Lorenz alternative readouts

(N=100) To test whether the Mackey–Glass vs MC trade-off might be mitigated by stronger readouts, we evaluated alternative readouts on a standard Lorenz 1-step forecasting setup (deterministic transitions; N=100; 5 seeds) . A kernel ridge (RBF) readout did not overturn the observed trade-off (Table S5): performance varied by seed and the mean remained modest relative to linear baselines shown elsewhere.

Table S5: **Lorenz 1-step** (N=100

) : alternative readout summary (5 seeds; mean ± 95% CI).

Readout	R^2 (test)	NRMSE (test)
Kernel ridge (RBF)	0.247 [−0.002, 0.501]	0.847 [0.689, 1.001]

The linear ridge baseline (Figures in the main text) remains a reasonable reference; the kernel readout did not qualitatively change conclusions for this protocol. Heavier polynomial expansions were found to be resource-intensive under the current feature path and are deferred to future work.

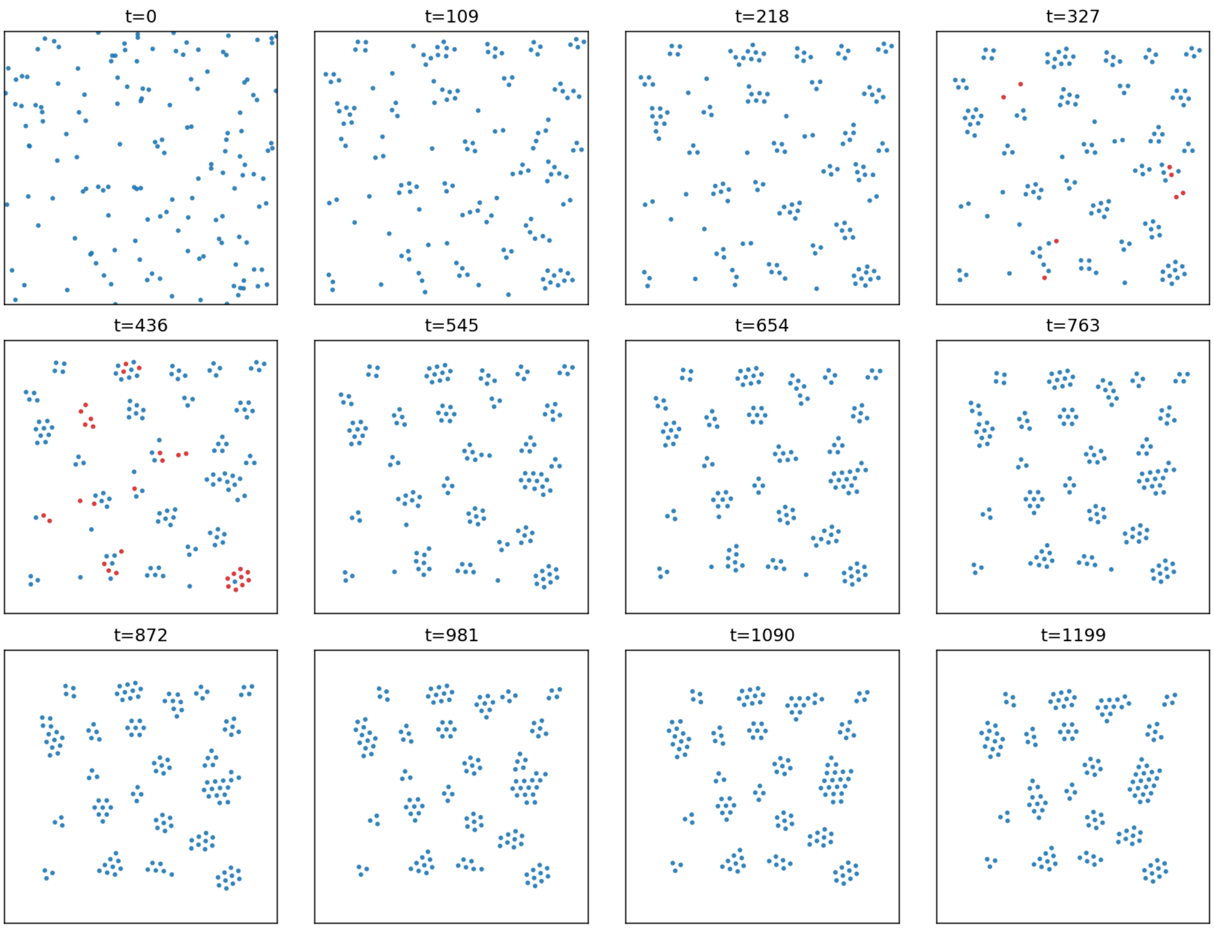

Figure S18: Illustrative brief-burst regime at N=150: mostly dispersed (state s_0) with intermittent, input-aligned excursions to state s_1 that align with higher MC.

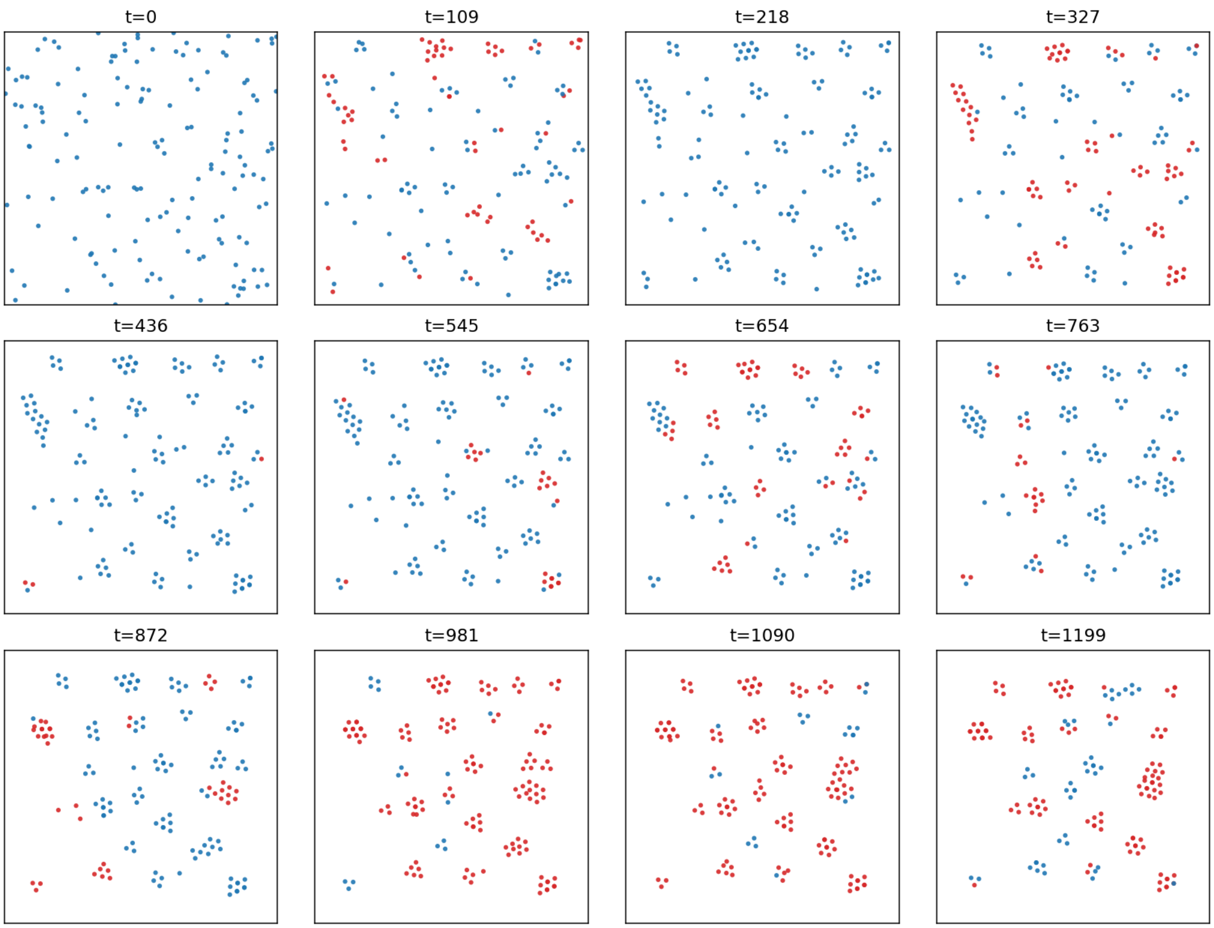

Figure S19: Illustrative higher clustered occupancy at N=150 with persistent red clusters; MC is lower than in the brief-burst example.

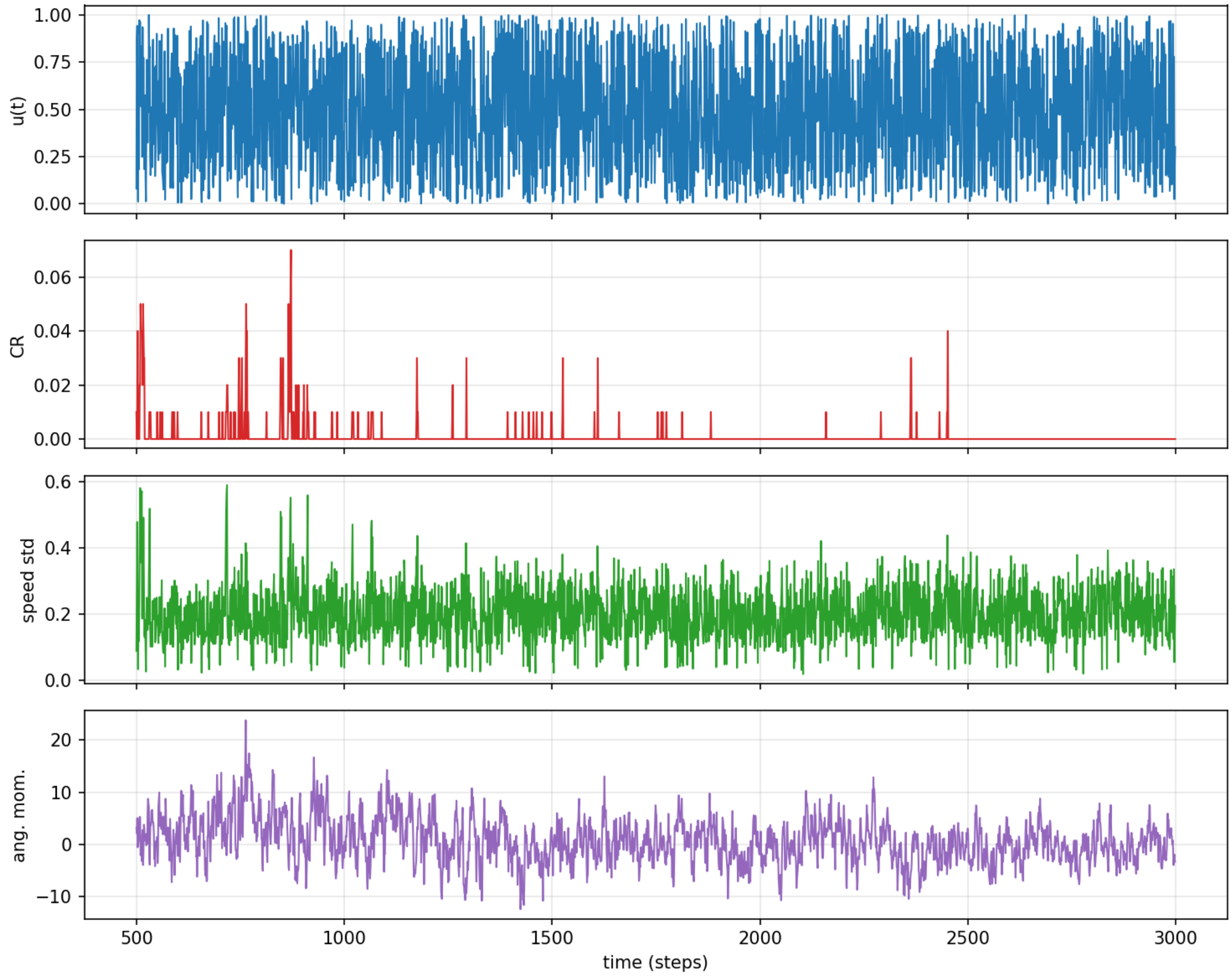

Figure S20: Representative post-washout segment in the brief-burst regime: input $u(t)$ and selected slow variables (clustering ratio f_{CR}, speed standard deviation f_{σ_v}, and angular momentum f_{AM}). Window starts after the initial transient. In this window f_{CR} is near-zero for most steps with rare excursions (mean $\approx$0.001; 95$^{\mathrm{th}}$ percentile $\approx$0.01; 99$^{\mathrm{th}}$ percentile $\approx$0.03; $f_{\mathrm{CR}} \geq 0.01$ on $\sim$5.4% of steps and $f_{\mathrm{CR}} \geq 0.03$ on $\sim$1.2%), consistent with brief bursts. (Interpreting f_{CR} as the fraction of agents in the clustered state, for this run N=100 these correspond to $\sim$0.1 agents on average and $\sim$3 agents at the 99$^{\mathrm{th}}$ percentile.)

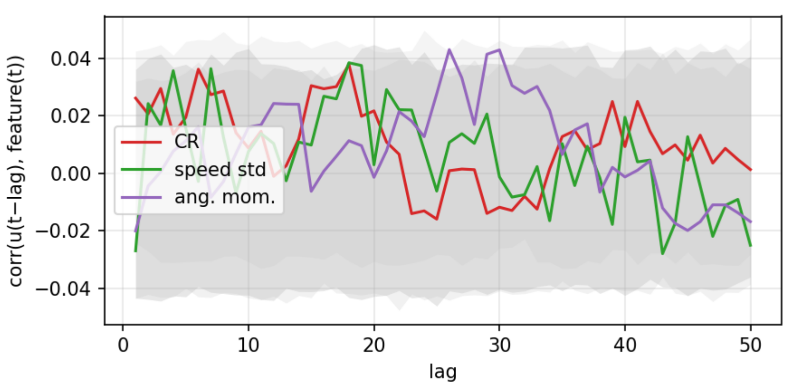

Figure S21: Normalized cross-correlation vs lag between standardized $u(t-\tau)$ and f_{CR}, f_{σ_v}, f_{AM} for the same run/window as Figure S20. Shaded bands show a circular-shift baseline (2.5–97.5th percentiles; n=500 shifts). Observed correlations are small ($|r| \lesssim 0.04$) and concentrated at small lags. We use xcorr as a compact alignment check rather than a causal claim.

S2.16 Parameter Tables

Table S6 lists the two-state operating point used unless otherwise stated.

Table S6: **Two-state operating point (values used unless otherwise stated)**

Parameter	Dispersed	Clustered
Outer radius	28.0	35.0
Alignment gain w_{align}	0.5	1.0
Cohesion gain w_{coh}	0.5	2.0
Separation gain w_{sep}	1.2	1.0
Shared constants: Inner radius = 15.0, α=0.4, $\lambda_s = 0.7$, speed clip ≈ 3.0, $dt \approx 0.1$.		

S2.17 Feature effective-rank diagnostic

Because the main scaling pipeline uses a degree-2 polynomial expansion of standardized features (54D), it is natural to ask whether the observed MC magnitudes and near-linear trends are consistent with the dimensionality actually explored by the observed feature process, rather than being an artifact of a trivially low-rank signal.

As a compact, non-claim diagnostic, we compute the effective rank of the post-washout feature time series as a function of swarm size N. Specifically, we compute (i) the participation-ratio effective rank of the standardized invariant 9 feature stream (9D), and (ii) the same quantity after a degree-2 polynomial expansion (54D; include_bias=false) with re-standardization. This diagnostic does *not* prove MC scaling from first principles; it only checks that the observed feature process is not collapsing to a near-1D manifold under the tested conditions.

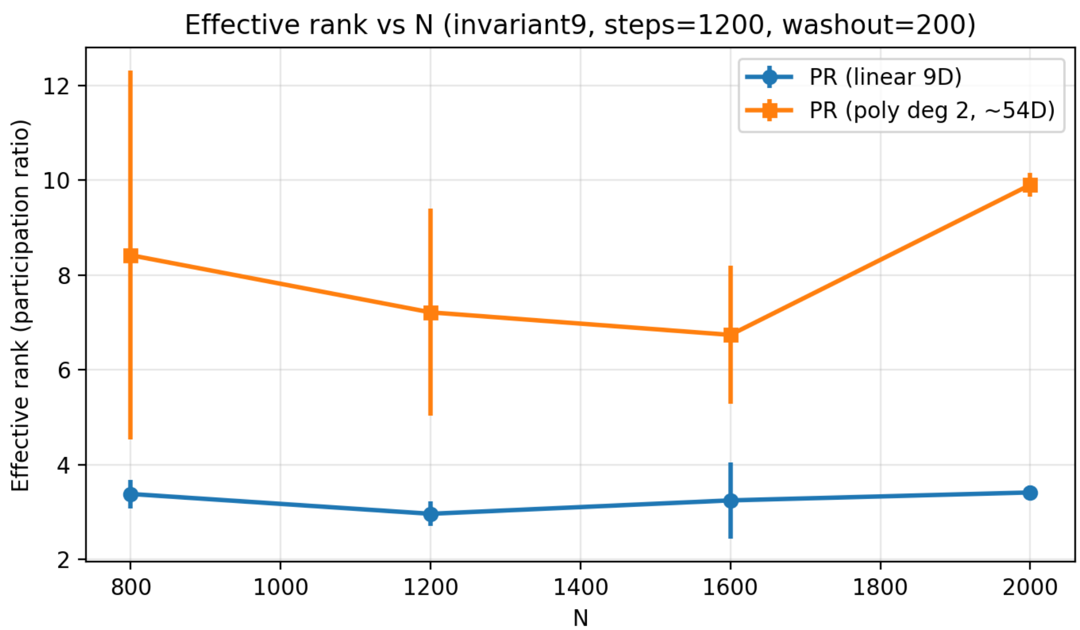

Figure S22: Effective rank vs swarm size.
We drive the two-state swarm with i.i.d. $u(t) \sim U(0, 1)$ (centred in the coupler), discard 200 steps of washout, and compute effective rank (participation ratio) on 1000 post-washout steps, averaged over seeds $\{0, 1, 2\}$. Blue: standardized invariant 9 features (9D). Orange: degree-2 polynomial expansion of standardized features (54D), with re-standardization. Across $N \in \{800, 1200, 1600, 2000\}$ the expanded feature stream has effective rank on the order of ∼7–10 (with entropy-rank ∼11–16), indicating the observed feature process is not trivially low-rank under the measured corridor.

S2.18 Mechanism ablations and mutual information

To address the "which ingredients are essential?" mechanism question, we ran two complementary ablation suites under the pure-MC protocol at a representative large population (N=800; steps=8000; washout=500; n=3 seeds) using the manuscript's input mode (temperature; 'mixed') and the base defaults (energy threshold = 3.0, hysteresis factor = 0.05).

Suite A: matched-runner (multistate feature path). This suite is matched to the paper's pure-MC runner pipeline and its feature-collection path (i.e., the same family of settings that reproduces the pure-MC magnitude at N=800). Conditions isolate: (a) *threshold-only* (no hysteresis), (b) *hysteresis-only with random coupling* (hysteresis retained but transitions driven by an i.i.d. random metric rather than local energy), (c) *full* (hysteresis + local-energy coupling), and (d) *naive_locked* (transitions locked; ‘num_states=2‘, ‘lock_states=true‘). Figure S23 summarizes total MC across these conditions.

As a complementary information-theoretic diagnostic, we estimate mutual information between delayed input and coarse macrostates $z(t)$ using a simple binned plug-in estimator. We report $I(u(t-\tau); z(t))$ for $z(t) \in \{f_{\mathrm{CR}}(t), f_{\sigma_v}(t), f_{\mathrm{AM}}(t)\}$ to match the slow-variable narrative, and include a circular-shift baseline band (Figure S24) to highlight the estimator's finite-sample bias floor. This is intended as a conservative "do these macrostates retain any measurable information about past input?" check, not as a first-principles derivation.

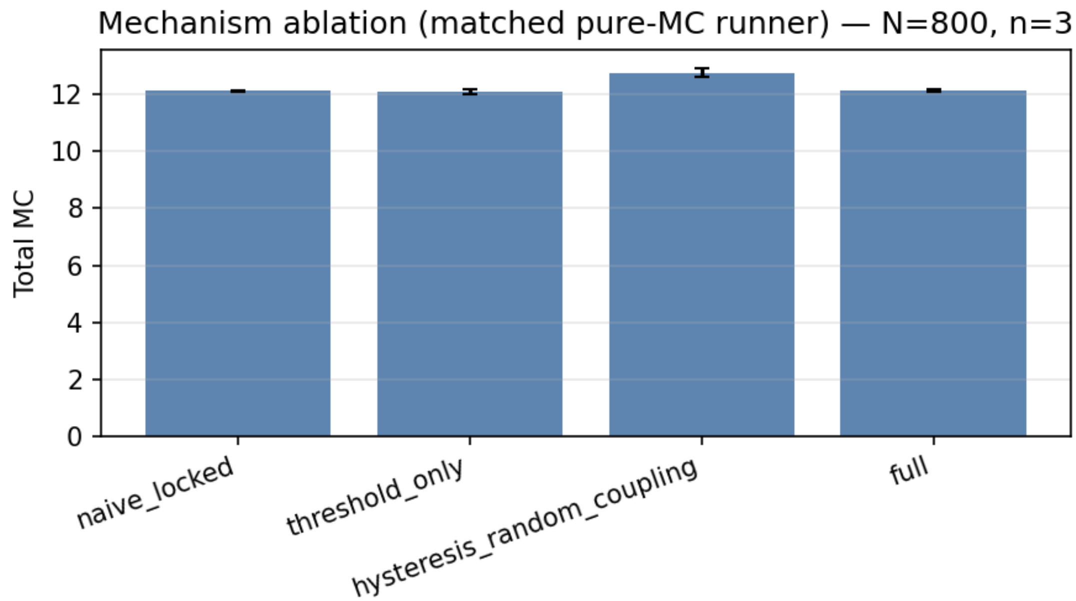

Figure S23: Suite A (matched-runner) targeted mechanism ablations under the pure-MC protocol at N=800 (steps=8000; washout=500; n=3 seeds). Bars show total MC (mean ± 95% CI across seeds). Conditions: naive_locked (transitions locked in a two-state model), threshold-only (no hysteresis), hysteresis-only with random coupling, and full (hysteresis + local-energy coupling). This suite reproduces the pure-MC magnitude of the paper's runner pipeline at N=800; differences between ingredients are modest at this n.

Suite B: state-agnostic invariant control (fixed observables). Because the definition of "single-state" in the paper corresponds to ‘num_states=1‘ (which implies a different feature path than ‘num_states=2‘), we also ran a control suite that enforces an identical, state-agnostic invariant feature map across all conditions (using ‘multi_state_features._single_state_features_list()‘ for both ‘num_states=1‘ and ‘num_states=2‘). This suite includes a true *single_state* (‘num_states=1‘), a *two_state_locked* control (‘num_states=2‘ with transitions locked), and the same threshold-only / hysteresis-random / full transition-rule variants. Under this fixed-observables control, total MC remains $\ll 1$ across conditions at N=800, indicating that the large pure-MC magnitude in Suite A is sensitive to the readout-visible representation available in the multistate pipeline.

References

[1] Mantas Lukoševičius. A practical guide to applying echo state networks. In Grégoire Montavon, Gene B. Orr, and Klaus-Robert Müller, editors, *Neural Networks: Tricks of the Trade (2nd ed.)*, volume 7700 of *Lecture Notes in Computer Science*, pages 659–686. Springer, 2012.

[2] Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.

[3] Chester Wringe, Martin Trefzer, and Susan Stepney. Reservoir computing benchmarks: a tutorial review and critique. *International Journal of Parallel, Emergent and Distributed Systems*, pages 1–39, 2025.

[4] Claudio Gallicchio and Alessio Micheli. Echo state property of deep reservoir computing networks. *Cognitive Computation*, 9:337 – 350, 2017.

Figure S24: Suite A (matched-runner) mutual information vs lag for the same ablation suite as Figure S23. Curves show the estimated $I(u(t-\tau); z(t))$ for $z(t) \in \{f_{\mathrm{CR}}(t), f_{\sigma_v}(t), f_{\mathrm{AM}}(t)\}$. Shaded bands indicate a circular-shift baseline (5th–95th percentiles; n=100 shifts) to highlight the estimator's finite-sample bias floor.

[5] Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, 2004.

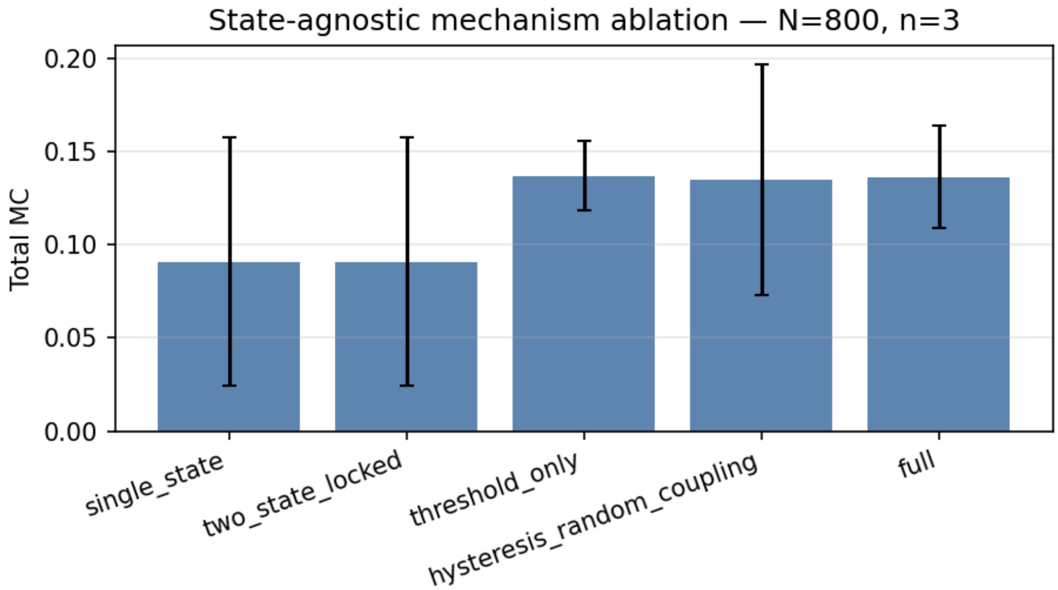

Figure S25: Suite B (state-agnostic invariant control) total MC under a fixed feature map shared across ‘num_states=1‘ and ‘num_states=2‘ (same N=800, steps=8000, washout=500, seeds $\{0, 1, 2\}$). Under this control, totals remain $\ll 1$ across conditions, and the ‘two_state_locked‘ control matches ‘single_state‘ within seed variability.

Figure S26: Suite B (state-agnostic invariant control) mutual information vs lag for the same conditions as Figure S25, with the same estimator and circular-shift baseline band.