CONTEXT-DEPENDENT PHASE RECONFIGURATION SEPARATES CONFIGURATION FROM INTRINSIC COMPUTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Recurrent systems often need to change their context-dependent representations while preserving useful computations. We study this problem in a periodic continuous-attractor model of hippocampal-to-entorhinal reconfiguration. A compatibility condition separates environment-dependent phase offsets from the spatial derivative that defines path integration. Structure-preserving phase translation follows a state-dependent tangent direction whose family lies in a fixed two-dimensional first-harmonic subspace per module. With recurrent dynamics frozen, a learned high-dimensional feedback interface expresses this predicted effective control geometry. Paired projection interventions support the functional role of the local tangent component, while post-capture interventions show that compatible feedback can remain part of ongoing closed-loop computation. An independent recurrent system dissociates target representation, instantaneous latent phase, and phase dynamics. In an external waypoint task, preserving dynamics provides a growing behavioral advantage after reconfiguration; this advantage persists after matching latent phase and external kinematics. Together, these results frame adaptation as computation-preserving reconfiguration, evaluated through both the state reached and its subsequent evolution.

1 INTRODUCTION

Biological and artificial recurrent systems often need to change the context they represent while preserving computations that remain useful across contexts. Navigation provides a concrete example. Hippocampal place-cell populations can remap across environments, while medial entorhinal grid populations can undergo coherent realignment while preserving structured relationships within local ensembles and modules (Fyhn et al., 2007; Stensola et al., 2012; Yoon et al., 2013). Such observations suggest a distinction between configuration—where a recurrent population represents the current state in a particular context—and computation—the dynamical rule by which that state is updated. For a path-integrating circuit, an environmental change may alter the phase alignment of the representation while leaving the self-motion-to-phase transformation useful across contexts.

Continuous-attractor models have long provided mechanisms for such state translation. In classical head-direction ring attractors, localized activity bumps are shifted by movement-dependent asymmetric interactions or by conjunctive rotation populations with fixed offset connectivity (Skaggs et al., 1995; Zhang, 1996). Zhang (1996) showed that an odd component proportional to the derivative of the symmetric recurrent interaction generates rigid bump translation, while fixed clockwise- and counterclockwise-biased pathways can realize the required asymmetry through movement-dependent mixing. Related mechanisms were extended to two-dimensional continuous-attractor models of grid-cell path integration, where velocity-dependent asymmetric drive translates population activity across the attractor manifold (Fuhs & Touretzky, 2006; Burak & Fiete, 2009). These results establish that fixed circuitry can support state-dependent translation; landmark and sensory inputs can additionally calibrate attractor phase.

Here we ask a complementary question. Suppose the recurrent dynamics implementing a useful computation are held fixed, while a context-dependent feedback pathway is allowed to reconfigure the represented state. Which target configurations are dynamically compatible with the original

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.

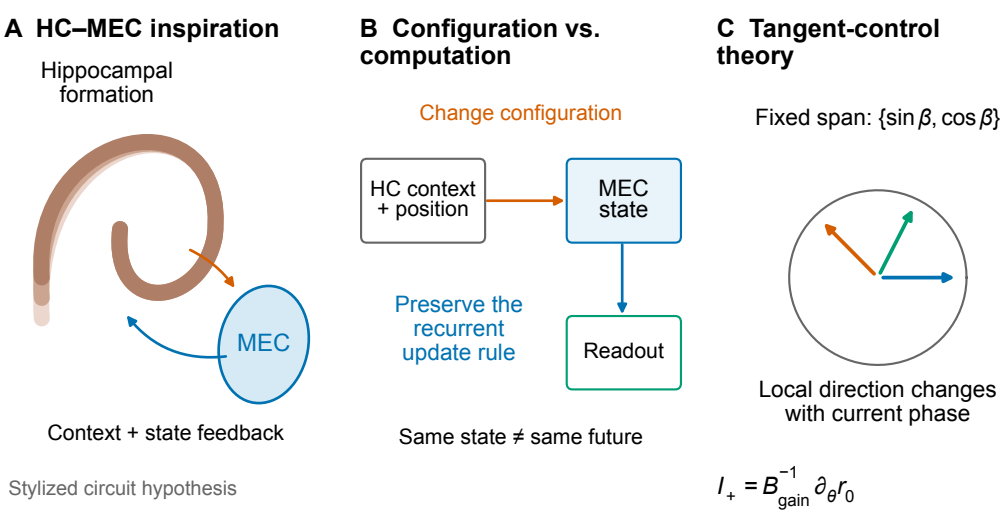


Figure 1: Computation-compatible reconfiguration. (A) HC–MEC circuit motivation. (B) Configuration changes while the recurrent update rule is preserved. (C) Phase-dependent translation directions lie in a fixed first-harmonic plane. Panels are conceptual or analytical illustrations.

computation, what control directions can reach them without disrupting the recurrent representation, and can a generic learned controller discover this structure without an explicitly prescribed shift circuit? This question also differs from recurrent-network studies in which context selects or modulates the computation itself (e.g., Mante et al., 2013; Beiran et al., 2023; Driscoll et al., 2024): our focus is on changing configuration while preserving the dynamical rule acting on it.

We address this problem first in a periodic HC-to-MEC model. Preserving the intrinsic path-integration rule constrains environment-dependent targets to retain the spatial derivative implemented by the recurrent circuit while allowing module-specific phase offsets to change. Structure-preserving reconfiguration follows the local attractor tangent; for the sinusoidal model, the family of these tangent directions lies in a fixed two-dimensional first-harmonic subspace, connecting the resulting control geometry to classical attractor-shift mechanisms. We then freeze MEC recurrent dynamics and ask whether functional learning organizes a high-dimensional hippocampal feedback interface into this theoretically predicted effective control geometry. Causal interventions distinguish selection of the global control repertoire from selection of the locally appropriate tangent direction and show that computation-compatible feedback can remain part of the closed-loop dynamics after reconfiguration. Finally, in an independent recurrent system with context-specific limit cycles, we dissociate representation, instantaneous state, and computation, and show that preserving the latter becomes increasingly consequential as subsequent behavior unfolds.

2 Compatibility and State-Dependent Tangent Control

We model each MEC module as a periodic continuous attractor with population phase θ_m. Its intrinsic dynamics implement $\dot{\theta}_m = k_m v(t)$, where $v(t) = \dot{x}(t)$ and k_m converts displacement into phase advance. An environment E specifies a target phase field $\theta_m^*(E, x)$.

2.1 Compatibility preserves the target derivative

For a fixed environment, define the tracking error as $e_m = \theta_m^*(E, x) - \theta_m$. If feedback contributes a phase-projected input I_{proj}, then $\dot{\theta}_m = k_m v + I_{\text{proj}}$, so $\dot{e}_m = (\partial_x \theta_m^* - k_m)v - I_{\text{proj}}$. Thus, matching the target at one instant does not ensure that it remains correct during motion. Preserving the intrinsic path-integration rule for arbitrary motion, without continuous phase forcing to compensate for a mismatched target derivative, requires $\partial_x \theta_m^* = k_m$. Integration gives $\theta_m^*(E, x) = k_m x + \phi_m^E$: context may change the offset ϕ_m^E, while the spatial metric k_m remains fixed. Appendix A expands the derivation and its scope.

2.2 Structure-preserving translation follows the attractor tangent

Let $r(\beta;\theta)$ denote the family of population patterns, with β indexing neural position around the attractor. A small translation satisfies $r(\beta;\theta+\delta\theta) = r(\beta;\theta) + (\partial_\theta r)\delta\theta + O(\delta\theta^2)$. The tangent $\partial_\theta r$ therefore changes phase while preserving the population pattern to first order; transverse components deform that pattern.

For the modeled bump and input port, the positive translation direction is $u_+(\beta;\theta) = 0.8\sin(\beta-\theta)$, with the coefficient set by the model's amplitude and normalization conventions. This direction changes with the current phase. Structure-preserving transport consequently requires feedback to depend jointly on the desired configuration and the current attractor state, rather than on the target alone.

Scaling a generic input changes both its translating and deforming components: stronger input can accelerate capture while increasing distortion, whereas weaker input can slow transport; selecting the tangent addresses this strength–distortion tradeoff through direction.

2.3 Fixed connectivity supports a rotating local direction

Expanding the tangent gives $u_+(\beta;\theta) = 0.8[\sin\beta\cos\theta - \cos\beta\sin\theta]$. All such directions lie in the fixed plane $\mathcal{U}_m = \mathrm{span}\{\sin\beta, \cos\beta\}$, although only one line in that plane is tangent at a given phase (Figure 1C). Fixed basis patterns with state-dependent coefficients can therefore implement the changing direction through fixed connectivity.

We express this feedback as $I_{\mathrm{HC}}(t) = G_\Theta(p^E(x_t), r_{\mathrm{MEC}}(t))$, where hippocampal activity supplies environment- and position-dependent information, MEC activity supplies the current state, and Θ contains fixed interface parameters. Across modules, feedback can select different target offsets and translate each module along its own tangent while retaining its path-integration scale. This predicts a two-dimensional effective repertoire per module and a state-dependent local action, without requiring a rank-two raw weight matrix. This factorization is closely related to classical ring-attractor shift mechanisms (Skaggs et al., 1995; Zhang, 1996); here it arises as the control geometry required by the reconfiguration problem and serves as a prediction for learned HC–MEC feedback.

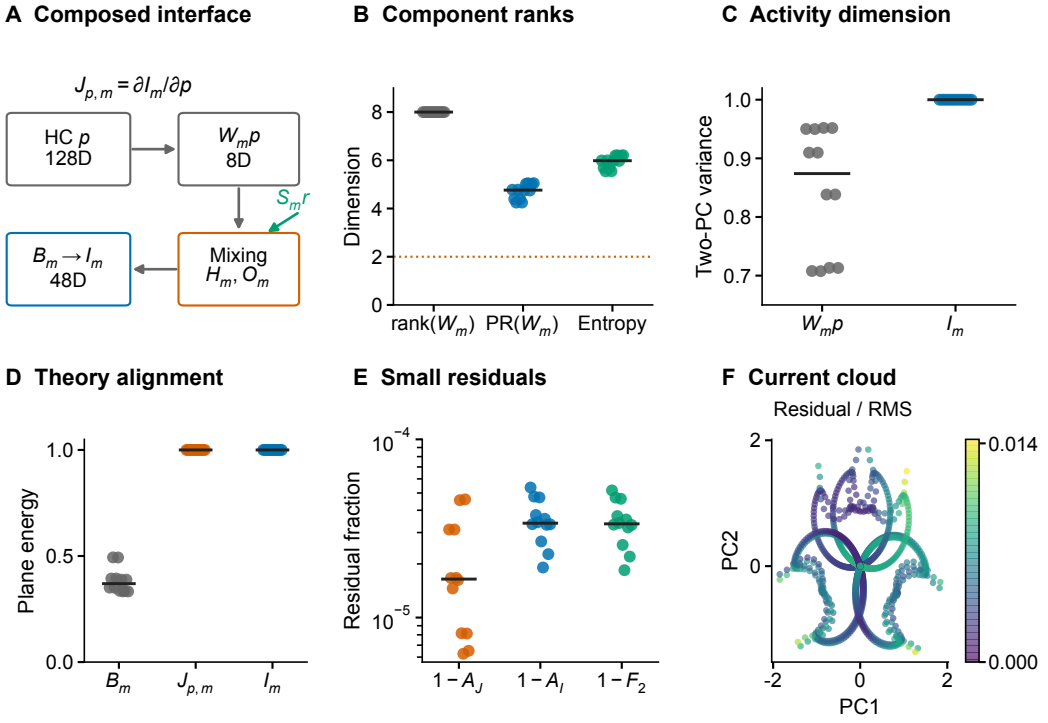
3 Learned feedback geometry and persistent closed-loop computation

3.1 Learned feedback expresses the predicted effective control plane

We froze the recurrent MEC dynamics and trained the hippocampal feedback interface to achieve target reconfiguration, preserve population structure, and maintain post-capture path integration. The training objective was $\mathcal{L}_{\mathrm{core}} = \lambda_T\mathcal{L}_T + \lambda_S\mathcal{L}_S + \lambda_{\mathrm{PI}}\mathcal{L}_{\mathrm{PI}}$, where L_T is the mean terminal circular phase error $1-\cos(\theta-\theta^\star)$, L_S is the time-weighted smooth maximum (log-sum-exp temperature 40) of the decoded-phase, gain-optimized bump deformation, and L_{PI} is the mean squared unwrapped phase-increment error relative to $k_m\Delta x$ from the first target-and-shape capture sustained for 2 model-time units to the end of that environment segment (requiring at least 2 post-capture units); the fixed normalization constants were estimated from development data, and the two environment segments were weighted equally. The interface combined hippocampal and current MEC signals through successive state-dependent transformations (Figure 2A; Appendix B).

All evaluated held-out trajectories ultimately reached the target, recovered accurate post-capture path integration, and remained captured over long-horizon evaluation; this endpoint success did not preclude severe transient distortion during transport, examined below.

To distinguish available capacity from the map actually expressed, we first examined the hippocampal readout $W_m \in \mathbb{R}^{8\times 128}$. All 12 module blocks retained rank eight, with participation ratios 4.25–5.05 and entropy ranks 5.54–6.21 (Figure 2B). The first two principal components explained 0.708–0.952 of centered intermediate activity $W_m p$, compared with 0.999948–0.999982 of expressed-current variance (Figure 2C). Thus, the intermediate representation retained dimensions that were largely absent from the final current.



B–E: 12 module blocks; black bars: medians. F: seed 0, module 1.

Figure 2: Effective control aligns with the theoretical plane. (A) Composed HC-to-current pathway. (B) Raw readout rank, participation ratio, and entropy rank. (C) Two-PC variance fractions for intermediate activity and current. (D–E) Raw output-matrix, complete-Jacobian, and centered-current alignment with the analytical plane, with near-unity residuals shown on a log scale. (F) Representative centered-current PCA projection colored by theory-plane residual. Summary points are 12 module blocks; black lines indicate medians.

We then compared the raw neural output matrix B_m with the complete local HC-to-current Jacobian $J_{p,m} = \partial I_m / \partial p$, holding MEC state fixed. Let P_m project onto the predicted plane $\mathcal{U}_m$. Across the 12 module blocks, only 0.333–0.494 of the raw output matrix's squared Frobenius norm lay in this plane. In contrast, the complete Jacobian's alignment, $\|P_m J_{p,m}\|_F^2 / \|J_{p,m}\|_F^2$, was 0.999954–0.999994.

The same plane captured 0.999946–0.999981 of centered current energy along sampled trajectories (Figure 2D–E). Figure 2F visualizes a representative current trajectory in its first two principal components, colored by residual from the analytical plane. PCA concentration measures dimensionality, whereas projection onto the analytical plane tests geometric alignment. Together, these measurements locate the predicted structure in the composed, state-dependent map despite broader raw component capacity. The 12 module blocks are nested within the checkpoint collection, not 12 independent training seeds (Appendix B).

3.2 Local tangent action supports transport and ongoing feedback

The global two-dimensional repertoire contains the appropriate directions across phases, but does not specify the correct local one-dimensional action. At each phase, we decompose the input as $I_m = a_t u_{\tan} + a_{\text{other}} u_{\text{other}} + I_\perp$, where u_{other} is the within-plane direction orthogonal to the tangent and $I_\perp$ lies outside the plane. In the D2 module-1 validation cohort, every examined trajectory instance developed a transient relative bump-distortion peak $D_{\text{rel}} > 0.6$, despite eventual capture. Yet the median trajectory-wise 95th percentile of $\|I_\perp\| / \|I_m\|$ was only 0.13% (Figure 3A). The

final capture criterion and transport-phase structural fidelity therefore describe different aspects of controller performance.

The main directional abnormality occurred within the plane. Two model-time units before each distortion peak ("Peak − 2" in Figure 3A), $|a_{\mathrm{other}}|/\|I_m\|$ was approximately 0.79–0.81, while phase remained sufficiently well defined for directional analysis. Let ρ_1 denote normalized first-harmonic amplitude: it fell to approximately 0.19–0.20 one unit before the peak and a median near 0.04 at the peak. Figure 3B shades $\rho_1 < 0.2$, where phase inferred from the harmonic argument is unreliable. Consequently, the pre-peak measurements, rather than directions estimated at collapse, support the local-direction diagnosis.

Within trajectories, non-tangent magnitude $|a_{\mathrm{other}}|$ correlated with concurrent deformation at approximately 0.85 and deformation one time unit later at 0.90–0.91, whereas total current magnitude correlated negatively with deformation, at approximately -0.71 (median correlations). Where phase was reliable, the tangent coefficient still pointed along the shortest circular route to the target. These observations implicate an additional within-plane component rather than a reversed tangent command or a simple contemporaneous increase in total strength. They remain descriptive because feedback and state co-evolve; they motivate, but do not replace, interventions on the local direction (Appendix C).

From identical saved MEC states in the C-checkpoint cohort, we compared the original feedback with feedback retaining only the tangent component or removing that component. Across 16 selected starts with substantial transport distortion, tangent-only feedback reduced median maximum relative distortion from 0.5036 to 0.2514, while maintaining substantial target progress with median terminal error approximately 0.078. Removing the tangent increased median maximum distortion to 0.6464 and impaired target acquisition (Figure 3C–D).

Feedback also contributed after reconfiguration. In paired interventions from 36 settled starts, removing hippocampal feedback degraded subsequent phase integration, whereas retaining tangent-compatible feedback preserved performance more effectively (Figure 3E). Hippocampal feedback can therefore remain part of the closed-loop computation after target capture. Hippocampal inactivation also disrupts grid firing biologically (Bonnevie et al., 2013), motivating the study of ongoing drive without establishing the specific tangent mechanism tested here. Its successful coexistence with the recurrent core depends on preserving the required phase evolution. The target-derivative condition in Section 2.1 identifies the configuration family compatible with the intrinsic metric; the post-capture experiment evaluates how the realized feedback interacts with that computation.

4 Representation, state, and computation dissociate beyond MEC

Does the distinction between configuration and computation depend on the MEC geometry? We tested an independent frozen recurrent system with $N = 128$ units and five context-specific limit cycles sharing the intrinsic phase rule $\dot{\varphi} = \omega$. A learned controller rerouted activity between cycles. This system did not impose the MEC attractor or Fourier control geometry; architecture and training details are collected in Appendix D.

We compared three objectives: target-only control required the correct target manifold; target-plus-phase control additionally constrained instantaneous latent phase throughout the trajectory; full control additionally constrained phase dynamics and included a weak post-reconfiguration control penalty. These objectives separately test representation, state, and computation (Figure 4A–D).

Operationally, representation success tests attainment of the target manifold; phase preservation tests $\varphi(t)$ against its reference $\varphi^*(t) = \varphi_0 + \omega t$; computation recovery tests preservation of the intrinsic rate $\dot{\varphi} = \omega$. Writing $y = W(r - b)$, $\varphi_c = \mathrm{atan2}(y_{c,2}, y_{c,1})$ and $\hat{\rho}_c = |y_c|$ for the core's linear readout, the preregistered criteria are: arrival, $d_{\mathrm{surr}} < 0.1$ with

$$d_{\mathrm{surr}}^2 = \sum_{l \neq c} |y_l|^2 + (|y_c| - \rho_c)^2 + \|(I - P_S)(r - b)\|^2 ,$$

held persistently; phase preservation, $|E_{\mathrm{comp}}| < \pi/4$ with

$$E_{\mathrm{comp}}(T) = \mathrm{wrap}\left[\varphi_{c_2}(T) - \varphi_0 - \omega T\right];$$

computation recovery, post-switch RMSE of $|\dot{\varphi}_{c_2}/\omega - 1| < 0.10$ restricted to samples where the target channel carries amplitude; and release, $\|I\|/\|F_{\mathrm{core}}\| < 0.05$, also persistent. Post-switch

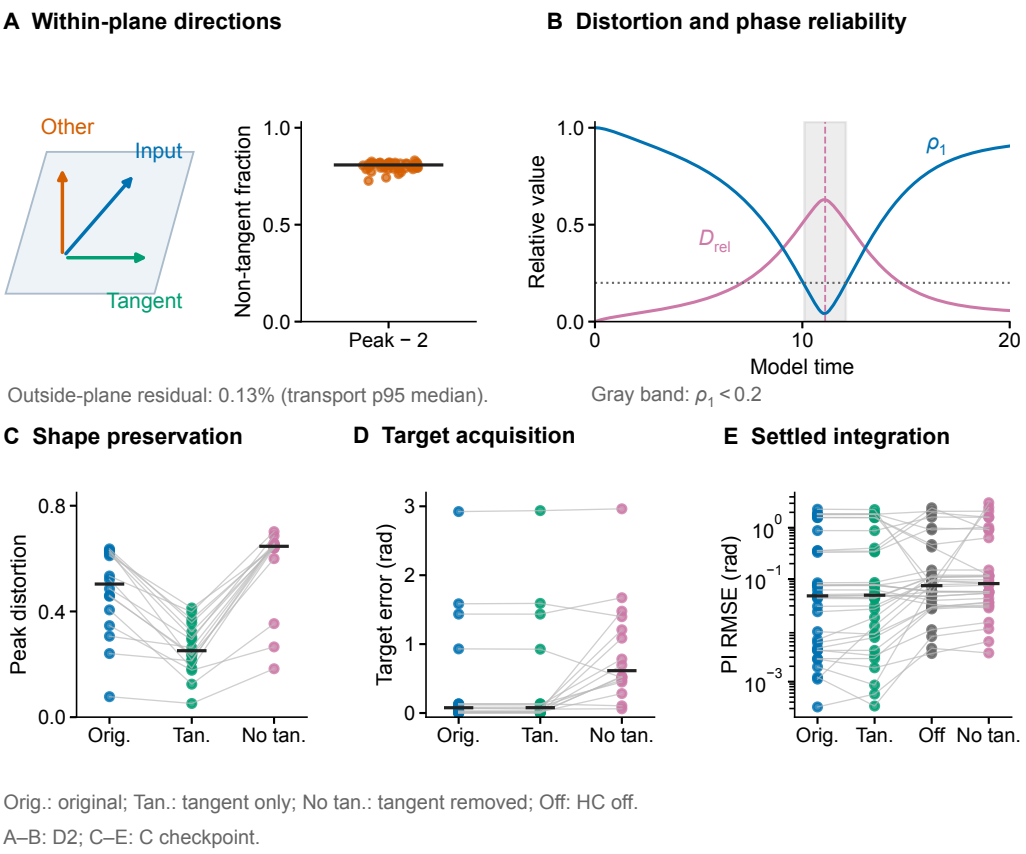


Figure 3: Within-plane direction links transport distortion to feedback interventions. (A) Non-tangent input fraction two model-time units before peak distortion in 48 saved module-1 trajectory copies from four D2 checkpoints; the annotation summarizes the transport out-of-plane fraction. Copies are not independent seeds. (B) Relative distortion D_{rel} and normalized first-harmonic amplitude ρ_1; gray shading marks unreliable phase estimates at $\rho_1 < 0.2$. (C–D) Tangent projections from 16 paired C-checkpoint starts. (E) Post-capture interventions from 36 paired settled starts. Lines connect paired starts; black lines indicate medians.

quantities are evaluated on $t \in [t_{\mathrm{switch}}, 4.0\ \mathrm{s}]$, with $t_{\mathrm{switch}} \sim U[0.6, 1.5]$ s drawn per trajectory. Per run we take the fraction of the 24 held-out trajectories satisfying each criterion, then the median and interquartile range over the 24 seeds, conditions paired by seed. These thresholds were fixed before production training and were not changed afterwards.

Across the evaluation summarized over 24 seeds, target-only control reached the target manifold in every trajectory but had computation recovery of 0.00. Target-plus-phase control preserved phase in 0.92 of trajectories, yet its phase-rate error was approximately three times that of full control (0.172 versus 0.054), and computation recovery was only 0.21. Correct target representation therefore did not ensure correct latent state, and accurate instantaneous phase did not ensure the intrinsic phase-update rule.

Removing the full controller's post-reconfiguration penalty preserved the reported representation, phase, and computation criteria while increasing residual control (Appendix D). This complements the HC–MEC result: persistent controller activity and recovered computation can coexist. The recovery criterion remains thresholded and does not imply identical trajectories over arbitrary horizons, which motivates the behavioral evaluation below.

Two interventions further connect the independent system to the control argument. Replacing learned feedback with a phase-averaged, transition-specific constant vector preserved mean input but removed

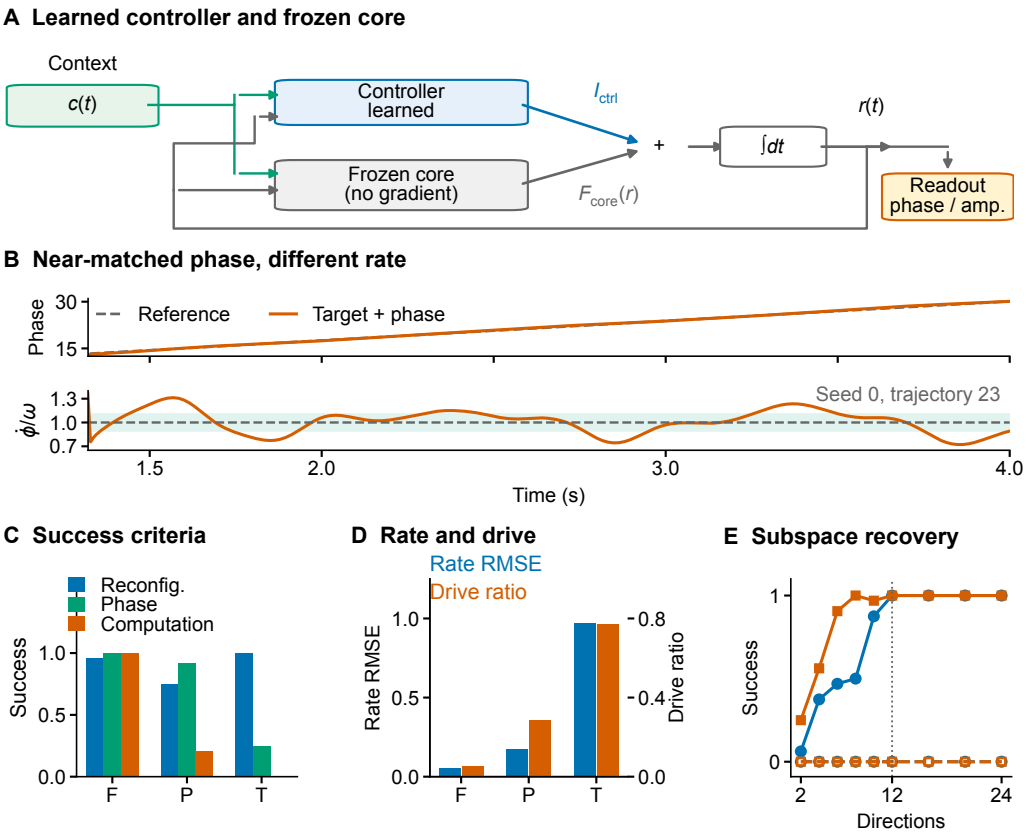
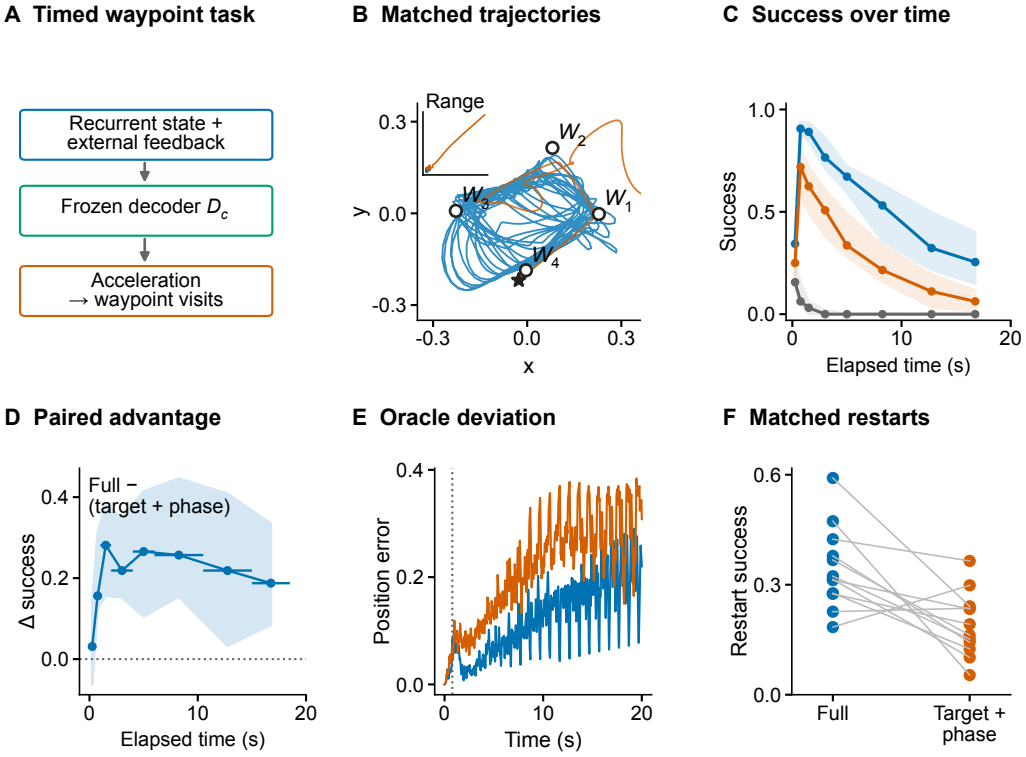


Figure 4: Phase accuracy and phase dynamics can dissociate. (A) Learned feedback acts on a frozen recurrent core. (B) A target-plus-phase trajectory tracks reference phase while its normalized instantaneous rate deviates from one; shading marks $\pm 10\%$. (C) Reconfiguration, phase, and computation success. (D) Post-switch rate RMSE and final control-to-core drive ratio, on separate axes. Panels C–D summarize 24 seeds. (E) Learned versus random output-subspace interventions across eight seeds; 12 is the smallest tested retained dimension yielding both learned-subspace success rates of 1.0.

dependence on current latent state, and failed to reproduce learned performance. Retaining learned output directions preserved function more effectively than a random subspace of equal dimension: across eight seeds, 12 was the smallest tested dimension at which learned-subspace reconfiguration and computation success both reached 1.0 (Figure 4E). Thus, both state dependence and the identity of control directions mattered; low dimensionality alone was insufficient.

5 Preserved dynamics support sustained behavior

We coupled the independent recurrent system to a two-dimensional waypoint task to test whether dynamical differences matter beyond the latent representation. Frozen downstream decoders converted recurrent activity and external state into point-mass acceleration, with each latent phase cycle corresponding to four waypoints. Decoders were trained on clean intrinsic trajectories and then frozen; held-out R^2 was approximately 0.997. With the correct intrinsic trajectory, the oracle achieved waypoint success 1.00, establishing that the downstream system could solve the task. Task configuration and training details are in Appendix E.



Blue: full; orange: target + phase; gray: target. Star: shared trajectory start.

C–D: 24 seeds, median and IQR. F: 12 seeds, 89 phase/external-state-matched restarts.

Figure 5: Dynamical preservation benefits sustained behavior. (A) Frozen decoders drive a timed waypoint task. (B–E) Trajectories, time-resolved success, paired differences, and oracle deviations; success curves summarize 24 seeds with medians and interquartile ranges. (F) Phase- and external-state-matched restarts across 12 seeds and 89 restarts. Full recurrent states are not matched.

Full control achieved overall waypoint success of 0.747, compared with 0.278 for target-plus-phase control, 0.023 for target-only control, and 0.051 without a controller. Full control thus improved performance substantially but remained well below the oracle. Target-only control performed worse than no control despite successful target-manifold attainment, showing that correct representation alone did not ensure useful subsequent behavior; the comparison does not separately identify residual input strength as the cause.

The full-versus-target-plus-phase advantage developed over time. During the first 0.5 s after switching, success was 0.344 versus 0.250; the paired difference was not significant ($p = 0.11$). From 0.5 s onward, full control showed a significant advantage. Between 1 and 2 s, the difference reached 0.281, with full control performing better in all 24 seeds. The gap persisted at longer horizons even as absolute performance declined in both conditions (Figure 5B–E). Thus, the behavioral cost of imperfect dynamics becomes more pronounced as the system generates a longer trajectory.

We then tested whether the advantage persisted after matching the current task-relevant state. States from the two conditions were selected with nearly identical latent phase and restarted from identical point-mass positions and velocities. Across 89 restarts from 12 seeds, waypoint success was approximately 0.319 for full control and 0.177 for target-plus-phase control, with full control better in 10 of 12 seeds (Figure 5F). Future behavior thus diverged despite matching phase and external kinematics. Complete recurrent states remained condition specific, so this comparison does not isolate a vector-field effect at an identical full neural state.

Together, the time course and restart experiments connect phase-dynamics preservation to sustained task performance, beyond target attainment or matching the measured present state. Passing the binary

computation criterion was still insufficient to predict equivalent long-horizon behavior: full_norel and mild_reg achieved waypoint success of 0.311 and 0.346, respectively, compared with 0.747 for full control, while retaining greater residual control. These differences show that thresholded recovery does not exhaustively describe dynamical fidelity (Appendix E).

6 DISCUSSION

Reconfiguration in a recurrent system is not only a state-placement problem. The target representation, the instantaneous latent state, and the dynamical rule acting on that state are distinct objects: reaching the appropriate representation need not recover the correct latent state, and matching the current latent state need not recover its subsequent evolution. Our results therefore frame reconfiguration as a problem of preserving computation while changing configuration. This view also clarifies the role of persistent feedback: the feedback structure allows feedback to remain part of the closed-loop computation over long timescales, and whether this long-term coexistence succeeds depends on whether the required dynamics are preserved.

The first-harmonic plane and local tangent direction are specific to the sinusoidal MEC model. Their fixed-basis, state-dependent realization is closely related to classical continuous-attractor shift mechanisms rather than a new general mechanism for translating bumps. Here, this geometry instead arises as a prediction from the requirement of computation-compatible reconfiguration and can therefore be tested against a learned feedback pathway. Functionally trained feedback organized high-dimensional interface components into an effective control geometry aligned with the predicted plane, and causal interventions confirmed the functional role of the local tangent direction. The broader result is not a particular two-dimensional basis, but a structured compatibility between external control and the recurrent dynamics whose computation is to be preserved. The independent recurrent system supports this distinction without imposing the Hippocampus-MEC geometry.

These results suggest that adaptation in recurrent systems should be evaluated through what happens after the target state is reached, not only through endpoint representation or instantaneous latent accuracy. This perspective is relevant to recurrent models in machine learning, where adaptation is often assessed by endpoint accuracy or representational similarity. Prior work has shown that computation in neural populations is carried by dynamics rather than static representations (Vyas et al., 2020), and that connectivity structure shapes the dynamics and computations of low-rank recurrent networks (Mastrogiuseppe & Ostojic, 2018). Our results extend this view by showing that preserving the dynamical rule is a distinct requirement from reaching the correct state. The present evidence comes from constructed recurrent systems; the restart experiments do not match complete recurrent states, and thresholded computation recovery does not exhaustively characterize long-horizon trajectories. Within these limits, the experiments show that small differences in recurrent evolution can accumulate into substantial behavioral differences, making dynamical fidelity an important criterion for successful reconfiguration. These results frame reconfiguration as a problem of computation-preserving adaptation, evaluated through both the state reached and its subsequent evolution.

AI USE STATEMENT

AI tools, including OpenAI Codex and GLM Zcode, were used to assist with writing and polishing the manuscript, information retrieval and discovery, developing proofs of mathematical claims, research ideation, and research execution, including coding and computational experiments. All AI-assisted outputs were reviewed and verified by the authors, who take responsibility for the final content.

REPRODUCIBILITY STATEMENT

All code used in this study has been uploaded as supplementary material.

REFERENCES

Manuel Beiran, Nicolas Meirhaeghe, Hansem Sohn, Mehrdad Jazayeri, and Srdjan Ostojic. Parametric control of flexible timing through low-dimensional neural manifolds. *Neuron*, 111(5):739–753.e8, 2023. doi: 10.1016/j.neuron.2022.12.016. URL https://doi.org/10.1016/j.neuron.2022.12.016.

Tora Bonnevie, Benjamin Dunn, Marianne Fyhn, Torkel Hafting, Dori Derdikman, John L. Kubie, Yasser Roudi, Edvard I. Moser, and May-Britt Moser. Grid cells require excitatory drive from the hippocampus. *Nature Neuroscience*, 16(3):309–317, 2013. doi: 10.1038/nn.3311. URL https://doi.org/10.1038/nn.3311.

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. URL https://doi.org/10.1371/journal.pcbi.1000291.

Laura N. Driscoll, Krishna V. Shenoy, and David Sussillo. Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. *Nature Neuroscience*, 27:1349–1363, 2024. doi: 10.1038/s41593-024-01668-6. URL https://doi.org/10.1038/s41593-024-01668-6.

Mark C. Fuhs and David S. Touretzky. A spin glass model of path integration in rat medial entorhinal cortex. *Journal of Neuroscience*, 26(16):4266–4276, 2006. doi: 10.1523/JNEUROSCI.4353-05.2006. URL https://doi.org/10.1523/JNEUROSCI.4353-05.2006.

Marianne Fyhn, Torkel Hafting, Alessandro Treves, May-Britt Moser, and Edvard I. Moser. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, 446:190–194, 2007. doi: 10.1038/nature05601. URL https://doi.org/10.1038/nature05601.

Valerio Mante, David Sussillo, Krishna V. Shenoy, and William T. Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503:78–84, 2013. doi: 10.1038/nature12742. URL https://doi.org/10.1038/nature12742.

Francesca Mastrogiuseppe and Srdjan Ostojic. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, 99(3):609–623.e29, 2018. doi: 10.1016/j.neuron.2018.07.003. URL https://doi.org/10.1016/j.neuron.2018.07.003.

William E. Skaggs, James J. Knierim, Hemant S. Kudrimoti, and Bruce L. McNaughton. A model of the neural basis of the rat's sense of direction. In Gerald Tesauro, David S. Touretzky, and Todd K. Leen (eds.), *Advances in Neural Information Processing Systems*, volume 7, pp. 173–180. MIT Press, 1995. URL https://proceedings.neurips.cc/paper/1994/hash/024d7f84fff11dd7e8d9c510137a2381-Abstract.html.

Hanne Stensola, Tor Stensola, Trygve Solstad, Kristian Frøland, May-Britt Moser, and Edvard I. Moser. The entorhinal grid map is discretized. *Nature*, 492:72–78, 2012. doi: 10.1038/nature11649. URL https://doi.org/10.1038/nature11649.

Saurabh Vyas, Matthew D. Golub, David Sussillo, and Krishna V. Shenoy. Computation through neural population dynamics. *Annual Review of Neuroscience*, 43:249–275, 2020. doi: 10.1146/annurev-neuro-092619-094115. URL https://doi.org/10.1146/annurev-neuro-092619-094115.

KiJung Yoon, Michael A. Buice, Caswell Barry, Robin Hayman, Neil Burgess, and Ila R. Fiete. Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature Neuroscience*, 16(8):1077–1084, 2013. doi: 10.1038/nn.3450. URL https://doi.org/10.1038/nn.3450.

Kechen Zhang. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience*, 16(6):2112–2126, 1996. doi: 10.1523/JNEUROSCI.16-06-02112.1996. URL https://doi.org/10.1523/JNEUROSCI.16-06-02112.1996.

APPENDIX

This appendix expands the derivations, implementation details, diagnostics, and additional experiments summarized in the main text.

A COMPATIBILITY AND CONTROL GEOMETRY

A.1 TRACKING-ERROR DERIVATION

For a fixed environment E, module m represents a phase $\theta_m(t)$ and has target field $\theta^*_m(E, x(t))$. The physical trajectory satisfies $\dot{x}(t) = v(t)$. With phase-projected feedback $I_{\mathrm{proj}}(t)$, the phase equation is $\dot{\theta}_m(t) = k_m v(t) + I_{\mathrm{proj}}(t)$. The error $e_m(t) = \theta^*_m(E, x(t)) - \theta_m(t)$ therefore obeys $\dot{e}_m(t) = \partial_x \theta^*_m(E, x(t))v(t) - \dot{\theta}_m(t) = (\partial_x \theta^*_m(E, x(t)) - k_m)v(t) - I_{\mathrm{proj}}(t)$.

This expression separates instantaneous alignment from compatibility with the recurrent update rule. Even when $e_m(t_0) = 0$, a mismatch between $\partial_x \theta^*_m$ and k_m regenerates error during motion unless an additional phase input compensates for that mismatch. Requiring the target field to remain compatible with intrinsic path integration for arbitrary motion, without such continuous compensation, gives $\partial_x \theta^*_m = k_m$. Integrating with respect to x yields $\theta^*_m(E, x) = k_m x + \phi^E_m$. A context switch changes ϕ^E_m while preserving k_m.

The derivative is taken along motion within a fixed environment; it does not describe the transient context switch itself. The argument identifies the target family compatible with the intrinsic spatial metric. It does not imply that the entire neural feedback vector must vanish after capture. Persistent feedback must instead be assessed through the phase evolution of the resulting closed-loop system. The post-capture interventions in Appendix C address this realized interaction experimentally.

A.2 TANGENT TRANSLATION AND TRANSVERSE DEFORMATION

The family of valid population profiles is $r(\beta; \theta)$, where β indexes neural position and θ is population phase. Taylor expansion gives $r(\beta; \theta + \delta\theta) = r(\beta; \theta) + (\partial r/\partial\theta)\delta\theta + O(\delta\theta^2)$. The derivative is tangent to the translated-pattern manifold. To first order, perturbations along this direction change phase, whereas transverse perturbations deform the population profile.

A generic perturbation can be separated into tangent and transverse components, $u = u_{\parallel} + u_{\perp}$. Increasing generic input strength can increase both target progress and deformation; decreasing it can reduce both deformation and transport speed. Tangent selection changes the orientation of control and addresses the geometric source of this tradeoff.

A.3 FIXED BASIS AND STATE-DEPENDENT COMPOSITION

The identity $u_+(\beta; \theta) = 0.8[\sin\beta\cos\theta - \cos\beta\sin\theta]$ expresses the phase-dependent tangent through two fixed patterns. The global repertoire for each module is $\mathcal{U}_m = \mathrm{span}\{\sin\beta, \cos\beta\}$. Its coefficients depend on the current phase, while the appropriate sign and amount of transport depend on the desired configuration.

A fixed controller $G_\Theta(p^E(x_t), r_{\mathrm{MEC}}(t))$ can use hippocampal target information and the current MEC state to select these coefficients. Fixed connectivity therefore supports changing local directions without online rewiring. This implementation principle is related to the fixed asymmetric pathways of classical ring-attractor models (Skaggs et al., 1995; Zhang, 1996); the present factorization specifies a prediction for feedback in the reconfiguration task, rather than an independent invention of state-dependent shifting through fixed circuitry. In multiple modules, the desired offsets $\{\phi^E_m\}_{m=1}^M$ and local translation directions are module specific. The two-dimensional result applies to each modeled module, not to arbitrary recurrent systems or to the rank of every interface component.

B HC–MEC ARCHITECTURE, TRAINING, AND EFFECTIVE DIMENSIONALITY

B.1 FEEDBACK INTERFACE AND FUNCTIONAL OBJECTIVE

The MEC recurrent dynamics were frozen, and only the hippocampal feedback interface was trained. For each module, the controller received a hippocampal population state $p^E(x_t)$ and the current MEC activity $r_{\mathrm{MEC}}(t)$, and generated current $I_m(t) = G_{\Theta,m}(p^E(x_t), r_{\mathrm{MEC}}(t))$.

The interface contained successive transformations. A readout $W_m \in \mathbb{R}^{8\times 128}$ mapped hippocampal activity into eight hidden channels. A separate map S_m read MEC activity. State-dependent and bilinear interactions in H_m combined these signals; O_m produced six output coefficients; and $B_m \in \mathbb{R}^{48\times 6}$ mapped those coefficients to a 48-dimensional neural current. The intermediate activity $W_m p$ is therefore distinct from both the final current and the complete HC-to-current map.

The training objective was $\mathcal{L}_{\mathrm{core}} = \lambda_T \mathcal{L}_T + \lambda_S \mathcal{L}_S + \lambda_{\mathrm{PI}} \mathcal{L}_{\mathrm{PI}}$, requiring target reconfiguration, population-structure preservation, and post-reconfiguration path integration. Training imposed no rank-two constraint on the HC readout, no first-harmonic output basis, no explicit tangent alignment, and no restriction of the final current to the theoretical plane. This architectural statement does not establish independence from geometry in the loss or supervision: a structure-preservation loss may favor manifold-compatible directions. All evaluated held-out trajectories ultimately reached the target, recovered accurate post-capture path integration, and remained captured over long-horizon evaluation, despite the transient transport distortion described in Appendix C.

B.2 RAW COMPONENTS RETAIN MULTIDIMENSIONAL CAPACITY

The dimensionality analysis covered three modules at each of four trained checkpoints, giving 12 module blocks. Every hippocampal readout retained full row rank, $\mathrm{rank}(W_m) = 8$. Participation ratios ranged from 4.25 to 5.05 and entropy ranks from 5.54 to 6.21. For centered intermediate activity, the fraction of variance explained by the first two principal components was $F_2(W_m p) = 0.708$–0.952 (Figure 2B–C).

Let P_m be the orthogonal projection onto $\mathcal{U}_m$. The raw output-matrix alignment was $A_B = \|P_m B_m\|_F^2 / \|B_m\|_F^2 = 0.333$–$0.494$. Substantial output capacity therefore remained outside the theoretical plane. These measurements concern component capacity; they do not alone determine the geometry of the composed controller.

B.3 COMPLETE-MAP AND EXPRESSED-CURRENT MEASUREMENTS

The complete local HC-to-current Jacobian was $J_{p,m} = \partial I_m / \partial p \in \mathbb{R}^{48\times 128}$, evaluated with current MEC state fixed. It includes the HC readout, MEC-state pathway, nonlinear mixing, output coefficients, and neural output basis. Across the sampled legal combinations of environment, position, and MEC phase, its theory-plane alignment was $A_J = \|P_m J_{p,m}\|_F^2 / \|J_{p,m}\|_F^2 = 0.999954$–$0.999994$.

Currents sampled along reconfiguration trajectories were centered across sampled task states before PCA. Their first two principal components explained $F_2(I_m) = 0.999948$–0.999982 of variance, with participation ratios 1.940–1.998 and entropy ranks 1.970–2.000. Direct projection onto the analytical first-harmonic plane captured 0.999946–0.999981 of centered current energy.

These measures answer different questions. PCA concentration estimates dimensionality using directions fitted to the sampled current; projection measures alignment with a plane specified by the theory. The complete Jacobian additionally measures the local sensitivity to hippocampal input at a fixed recurrent state. Agreement among them supports an effective control geometry that is both low dimensional and theory aligned, while the raw component measurements show broader available capacity.

Figure 2B–C reports component dimensionality and intermediate-versus-current PCA concentration. Figure 2F shows a representative centered-current PCA projection colored by its residual from the theoretical plane. Black lines in the summary panels indicate medians across module blocks.

C LOCAL DIRECTION, DEFORMATION, AND CAUSAL INTERVENTIONS

C.1 DESCRIPTIVE ANALYSIS WITHIN THE FIRST-HARMONIC PLANE

At a given phase, write $I_m(t) = a_t(t)u_{\tan}(t) + a_{\text{other}}(t)u_{\text{other}}(t) + I_\perp(t)$. Here $u_{\tan} \propto \sin(\beta - \theta_m)$, u_{other} is the orthogonal direction within the first-harmonic plane, and $I_\perp$ lies outside that plane. The out-of-plane residual is therefore not the complete transverse component: the within-plane non-tangent direction is also transverse to local translation.

In the common module-1 validation analysis, the median trajectory-wise 95th percentile of relative out-of-plane input was approximately 0.13% of input norm, yet every examined trajectory instance had transient relative distortion exceeding $D_{\text{rel}} > 0.6$. Figure 3A summarizes 48 saved module-1 trajectory copies from four checkpoints.

Approximately two model-time units before peak deformation, the first-harmonic amplitude still supported meaningful phase-dependent direction estimates, while $|a_{\text{other}}|/\|I\|$ was approximately 0.79–0.81. One unit before the peak, this fraction remained approximately 0.79–0.82, while normalized first-harmonic amplitude had fallen to approximately 0.19–0.20. At the peak, median normalized amplitude was near 0.04, making phase estimates based on the harmonic's argument unreliable. Figure 3B illustrates this collapse and shades the interval with normalized first-harmonic amplitude $\rho_1 < 0.2$, where instantaneous phase is unreliable.

Where phase remained well defined, the tangent coefficient retained the sign associated with progress along the shortest circular direction to the target. The abnormality was accompanied by a large additional within-plane component, rather than a simple reversal of the tangent command. Within trajectories, $|a_{\text{other}}(t)|$ had median correlation approximately 0.85 with concurrent deformation and 0.90–0.91 with deformation one model-time unit later. Total current magnitude instead had median correlation approximately -0.71 with deformation.

These correlations are descriptive. In closed loop, non-tangent current may contribute to deformation, respond to deformation, or participate in both processes. Near harmonic collapse, phase-based direction estimates are additionally unreliable. This analysis motivates interventions but does not independently identify a causal direction effect.

C.2 PAIRED TRANSPORT INTERVENTIONS

The causal comparison started from identical saved MEC states and generated future trajectories with original learned feedback, tangent-only feedback, or feedback with its tangent component removed. Figure 3C–D uses 16 selected high-distortion starts from a C checkpoint. It is a different checkpoint family from the descriptive D2 analysis in Figure 3A–B.

Median maximum relative distortion was 0.5036 under original feedback, 0.2514 under tangent-only feedback, and 0.6464 after tangent removal. Tangent-only feedback maintained substantial progress, with median terminal target error approximately 0.078; removing the tangent impaired target acquisition. Connecting lines identify paired starts, and black lines show medians. The reported $D_{\text{rel}} > 0.6$ descriptive result and the 0.5036 original-feedback intervention median come from different checkpoint families and samples and should not be combined as one cohort.

C.3 FEEDBACK AFTER CAPTURE

Figure 3E compares original feedback, tangent-only feedback, hippocampal feedback off, and tangent-removed feedback from 36 paired settled starts. Removing hippocampal feedback degraded subsequent phase integration, while retaining tangent-compatible feedback preserved performance more effectively.

These results show that feedback can remain useful after capture in the evaluated closed-loop system. The interpretation concerns preservation of the required phase evolution by feedback interacting with the recurrent core. It does not establish indefinite stability, universal necessity of persistent feedback, or a requirement that all residual feedback be tangent at every state.

D INDEPENDENT RECURRENT SYSTEM: SETUP, OBJECTIVES, AND ABLATIONS

D.1 RECURRENT CORE AND CONTROLLER

The independent system used a frozen recurrent core with $N = 128$ units and five coexisting context-specific limit cycles $\{\mathcal{M}_c\}_{c=1}^{5}$. Each cycle carried latent phase φ with the same intrinsic rate, $\dot{\varphi} = \omega$. A learned controller received context and recurrent-state information and rerouted the state from one cycle to another. Core and controller outputs combined to update the recurrent state, from which phase and amplitude were read out (Figure 4A).

The controller was not supplied with grid-cell geometry, a first-harmonic control basis, a tangent-normal decomposition, phase equivariance, or a low-rank factorization. The experiment tests the representation/state/computation distinction in a different constructed system; it is not a test that the MEC-specific Fourier mechanism transfers unchanged.

D.2 TRAINING CONDITIONS AND EVALUATION

Representation denotes reaching the requested context-specific manifold. State denotes preserving the appropriate instantaneous latent phase. Computation denotes preserving its intrinsic phase evolution.

Condition	Reported training constraints
target-only	Reach the target context-specific manifold.
target-plus-phase	Reach the target manifold and constrain instantaneous latent phase throughout the trajectory; no phase-rate constraint.
full	Add a phase-dynamics constraint and a weak post-reconfiguration control penalty.
full_norel	Remove the post-reconfiguration control penalty from the full objective.

Figure 4C–D summarizes 24 seeds. Target-only control reached the target manifold in every evaluated trajectory, but computation recovery was 0.00 and final control-to-core drive ratio was $q_{\mathrm{final}} = 0.77$. Target-plus-phase control preserved phase in 0.92 of trajectories and recovered computation in 0.21, with phase-rate error 0.172 compared with 0.054 for full control.

The representative target-plus-phase trajectory in Figure 4B is seed 0, trajectory 23. Its phase follows the reference after the switch, while its normalized instantaneous rate departs from one. The shaded band is $\pm 10\%$ around the intrinsic rate.

D.3 RESIDUAL-CONTROL ABLATIONS

Removing the post-reconfiguration penalty in full_norel left representation, phase, and computation preservation at or above the full condition's reported level, while substantially increasing residual control. The mild_reg condition showed a similar pattern. These results separate nominal computation recovery from the amount of controller activity remaining after reconfiguration.

Passing a thresholded recovery criterion does not require identical trajectories or guarantee identical external performance. The behavioral consequences of these conditions are reported in Appendix E.4.

D.4 PHASE-AVERAGED FEEDBACK

At evaluation, the learned controller was replaced with a phase-averaged vector that preserved the mean transition-specific input while removing dependence on the current latent state. This constant control failed to reproduce the learned controller's performance. Thus, target context and mean input alone did not reproduce the function of state-dependent feedback in this system.

D.5 Learned and random output subspaces

Causal subspace analysis retained learned output directions and compared them with a randomly oriented subspace of equal dimension (Figure 4E). Task-relevant control was concentrated in approximately 12 dimensions. Across eight seeds, 12 was the smallest tested retained dimension at which both reconfiguration and computation success reached 1.0 in the learned subspace. A random subspace of matched dimension did not preserve this function.

This establishes functional specificity among the tested directions, rather than a universal minimal dimension. In particular, the 12-dimensional result should not be identified with the analytical two-dimensional repertoire of a MEC module.

E Waypoint task: setup, controls, and additional results

E.1 External task and frozen decoder

The frozen recurrent system drove a two-dimensional point mass. One full latent-phase turn corresponded to a cycle through four spatial waypoints at the required times. Fixed, context-specific downstream decoders mapped recurrent activity and the current point-mass state to acceleration commands. Phase was not supplied directly to the decoder. Decoder training used only clean intrinsic recurrent trajectories, after which decoders were frozen and were not adapted following a context switch (Figure 5A).

Held-out decoder performance was approximately $R^2 = 0.997$, with near-ceiling task performance on clean context-specific trajectories. An oracle using the same downstream system and the correct intrinsic latent trajectory achieved waypoint success 1.00. These controls establish that the frozen downstream system can express the task on appropriate recurrent trajectories.

E.2 Overall and time-resolved performance

Reported overall waypoint success was 0.747 for full control, 0.278 for target-plus-phase control, 0.023 for target-only control, and 0.051 without a controller. Target-only control's poorer performance than no control is consistent with its persistent input perturbing useful evolution after target attainment; this comparison alone does not isolate residual input magnitude as the cause.

The time-resolved analysis summarizes 24 seeds with eight trajectories per seed, as annotated in Figure 5C. That panel shows median success with interquartile ranges, and Figure 5D shows within-seed full-minus-target-plus-phase differences, again summarized by medians and interquartile ranges. Representative matched-episode trajectories and deviation from an oracle trajectory appear in Figure 5B and E.

During the first 0.5 s after switching, success was 0.344 for full control and 0.250 for target-plus-phase control, with a nonsignificant paired difference ($p = 0.11$). This result is not an equivalence test and does not establish identical early behavior.

From 0.5 s onward, full control showed a significant advantage. Between 1 and 2 s, the difference was 0.281, and full control performed better in all 24 seeds. The advantage persisted over longer horizons while absolute success declined in both conditions.

E.3 Phase- and external-state-matched restarts

Restart states were selected from full and target-plus-phase trajectories with nearly identical task-relevant latent phase. The two point masses were restarted at identical positions and velocities, while each recurrent system continued under its respective condition-specific dynamics. Figure 5F summarizes 89 restarts across 12 seeds.

Waypoint success was approximately 0.319 for full control and 0.177 for target-plus-phase control. Full control performed better in 10 of the 12 seed-level comparisons. These are 89 restarts nested within 12 seeds, not 89 independent training runs.

The matching controls measured latent phase and external kinematics, but the complete recurrent states remained different. Unmatched neural-state components could therefore contribute to subsequent divergence. The result supports an advantage persisting beyond phase and external-state matching; it does not uniquely attribute divergence to different vector fields evaluated from an identical full recurrent state.

E.4 Computation recovery and long-horizon behavior

The full_norel and mild_reg conditions met the same phase- and computation-preservation criteria as full control, but achieved waypoint success of 0.311 and 0.346, respectively, compared with 0.747 for full control. Both retained substantially greater residual control.

This comparison shows that binary recovery of the principal phase computation does not exhaustively characterize long-horizon task performance. Smaller closed-loop perturbations can remain relevant after a criterion is passed. The reported association with residual control should not be read as a separate causal isolation of residual-control magnitude.