STRUCTURED PHYSICAL PRIMITIVES FOR ROBUST NEURAL DECODING

Anonymous authors
Paper under double-blind review

ABSTRACT

The main challenge of neural decoding in practical applications, such as brain-computer interfaces (BCIs), lies in maintaining stable performance in the presence of noise and missing data. Current methods either learn dynamics directly from data or impose statistical constraints, making them prone to degradation when data quality declines. Physical motion, however, possesses intrinsic geometric structure with motion-specific primitives, such as translation and rotation in rodent planar motion, which are independent of neural data quality and offer a stable source of prior knowledge when neural evidence is unreliable. Motivated by this, we propose Structured Physical Primitives for Neural Decoding (MaPIC), a novel framework comprising two complementary components: a neural evidence encoder for observation likelihood and a physical primitive model for dynamic Bayesian priors. By maximizing the ELBO, MaPIC continuously integrates neural evidence with physical priors, where the KL divergence term serves as a structured regularizer guided by physical motion primitives. Notably, MaPIC learns the posterior uncertainty of each primitive via variational inference, enabling adaptive weighting among primitives. Our framework is causal and plug-and-play, requiring no modifications to the base architecture. Experiments on simulated dynamical systems and real rat hippocampal recordings demonstrate that even when 80% of neurons is randomly dropped, MaPIC surpasses state-of-the-art methods, significantly reducing position decoding error and correcting directional phase delays at turning points.

1 INTRODUCTION

Neural decoding infers continuous behavior from high-dimensional population activity, central to neuroscience and brain–computer interfaces. While contrastive methods like CEBRA embed neural activity for nonparametric decodingSchneider et al. (2023), they rely on training distribution regularities, so under noise or neuronal loss from electrode drift or surgical constraints, the learned manifold becomes unreliable and decoding collapses when robustness matters most.

Prior knowledge can help constrain the decoder. Some methods impose temporal smoothness by penalizing velocity or acceleration. Others use Bayesian filtering to combine neural predictions with motion estimates Wu et al. (2003); Gilja et al. (2012). DPAD instead learns neural and behavioral dynamics with a recurrent network Sani et al. (2024). These methods offer useful temporal constraints, but do not explicitly use the geometry of physical movement as a prior. When neural evidence weakens, they may produce physically implausible trajectories.

Physical movement has a stable geometric structure. Translation and rotation are basic primitives of planar locomotion, including movement on tracks and in open fields. These primitives remain valid when neural recordings are noisy or neurons are missing. We therefore encode them as structured priors in the decoding process.

We propose **MaPIC** (Motion-aware Physical-primitive Informed Correction) for robust neural decoding. A neural encoder provides observation evidence, while a physical primitive model defines dynamic priors over translation and rotation. MaPIC combines them by maximizing an evidence lower bound (ELBO). Its KL term regularizes inference toward physical motion, and variational inference estimates uncertainty for each primitive. This allows the model to adjust their influence as neural evidence changes. MaPIC is causal and works with existing neural embeddings without

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.

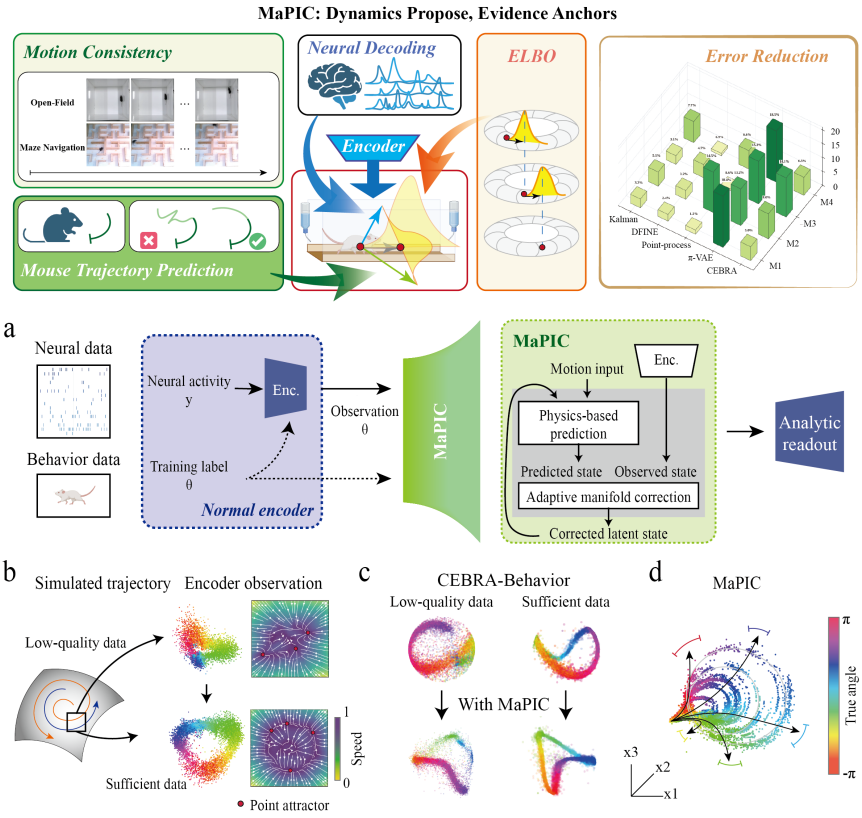


Figure 1: MaPIC combines neural evidence with physical motion priors to recover continuous, behavior-aligned latent trajectories under limited data.

changing the base encoder. It also resets recurrent states at trial boundaries to prevent information leakage.

We evaluate MaPIC on simulated systems and three neural datasets: anterior dorsal thalamic nucleus (ADn) head-direction, grid-cell, and hippocampal CA1 linear-track recordings. MaPIC improves decoding with limited supervision and sparse neural populations. Even with 80% neuronal dropout, it outperforms the baselines, reducing position error and directional phase delays at turns. These results support physical motion primitives as robust priors when neural evidence degrades.

Our contributions are as follows: **(1)** We show how noise and neuronal dropout weaken data-driven decoding, and identify the lack of explicit geometric motion priors in existing approaches. **(2)**We introduce MaPIC, which models translation and rotation as dynamic Bayesian priors. Its variational objective balances neural evidence with physical motion. **(3)** MaPIC is causal, interpretable, and compatible with existing neural encoders. Boundary-aware resets preserve trial structure. **(4)** In both the simulation system and real neural recordings, MaPIC significantly improves the decoding performance in both the basic model and the task, and still outperforms the best method even when 80% of the neurons are dropped out.

2 Related Work

Neural decoding has progressed along three largely separate lines: learning nonlinear neural representations, exploiting geometric structure in population activity, and performing probabilistic state estimation.

Neural representations and learned dynamics. Low-dimensional population models established that neural activity lies on a low-dimensional manifold Cunningham & Yu (2014); Churchland et al. (2012); Gallego et al. (2018); Vyas et al. (2020). Nonlinear representation methods, including identifiable π-VAE and contrastive CEBRA, learn embeddings that preserve behavioral information Zhou & Wei (2020); Schneider et al. (2023). Dynamical models such as LFADS and DPAD add temporal structure through recurrent or sequential architectures Pandarinath et al. (2018); Sani et al. (2021; 2024); Hurwitz et al. (2021). These approaches provide strong statistical constraints, but their dynamics are learned implicitly from data. When neural observations are noisy or a large fraction of neurons is missing, the learned dynamics lack a principled fallback, and decoding degrades.

Geometric structure and motion priors. Attractor theories and empirical ring or toroidal organization motivate structured behavioral state spaces Chaudhuri et al. (2019); Gardner et al. (2022); Khona & Fiete (2022). Manifold GPLVMs and probabilistic geometric PCA incorporate non-Euclidean structure into latent-variable modeling Jensen et al. (2020); Hsieh & Shanechi (2025). MARBLE and coherence objectives encourage geometrically organized representations Gosztolai et al. (2025); Gaukstad et al. (2026). These methods exploit geometry as a representational constraint, but they typically do not adapt the influence of geometric structure according to the reliability of neural evidence, nor do they explicitly model how motion primitives should be weighted when observations become unreliable.

Probabilistic state estimation and adaptive fusion. Kalman decoding and latent state-space models integrate motion predictions with neural observations Wu et al. (2003); Gilja et al. (2012); Aghagolzadeh & Truccolo (2016); Brandman et al. (2018). Variational filters and deep state-space models extend observation models and learning procedures Krishnan et al. (2015); Karl et al. (2017); Burkhart et al. (2020). Physics-informed variational models incorporate known dynamics into inference Zhong & Meidani (2023), while robust estimation addresses unreliable information Huber (1964); Masreliez & Martin (1977). However, classical filters rely on linear or Gaussian assumptions, treat motion prediction as generic smoothness, and do not explicitly encode physical primitives. They also do not target severe neuronal dropout, where the observation model itself becomes ill-defined.

In sum, existing methods either ignore the physical structure of behavior, reduce it to a generic smoothness prior, or encode it implicitly in black-box recurrent weights. None exploits movement geometry as a stable, interpretable prior, so when neural evidence degrades they lack a principled fallback. MaPIC addresses this by encoding motion-specific primitives explicitly on the state manifold, inferring their posterior uncertainty via variational inference, and fusing them with neural evidence through uncertainty-dependent assimilation.

3 METHODS

3.1 PROBLEM FORMULATION

Let $\mathcal{M}$ denote the behavioral state space, a d-dimensional smooth manifold specified by the physical structure of the task. We consider discrete time steps $t = 1, \ldots, T$. At each step, neural activity $y_t \in \mathbb{R}^{n_y}$ provides an observation of the behavioral state $x_t \in \mathcal{M}$ through a pointwise mapping

$$\tilde{x}_t = D(F(y_t)) \in \mathcal{M}, \qquad r_t = \rho(y_t) \in \mathbb{R}^{n_r}, \tag{1}$$

where $\tilde{x}_t$ is a noisy estimate of x_t, and r_t summarizes its reliability. Our objective is to infer a trajectory $\hat{x}_{1:T}$ that remains accurate when $\tilde{x}_t$ degrades due to recording noise or the loss of a large fraction of neurons.

The inductive bias we exploit is that behavioral motion on $\mathcal{M}$ is not arbitrary: it possesses intrinsic geometric structure with motion-specific primitives, i.e., elementary displacements along the local degrees of freedom of $\mathcal{M}$. Let $A_t \in \mathbb{R}^d$ denote the latent motion coordinates at time t, where each coordinate corresponds to one primitive. The state evolves on $\mathcal{M}$ according to

$$\bar{x}_t = \Phi_{\hat{x}_{t-1}}\left(\Delta t\, B_{\hat{x}_{t-1}} A_t\right) \in \mathcal{M}, \tag{2}$$

where $B_x : \mathbb{R}^d \to T_x\mathcal{M}$ is a local frame at x, $\Phi_x : T_x\mathcal{M} \to \mathcal{M}$ is the local retraction, and Δt is the sampling interval. This construction ensures that every motion proposal $\bar{x}_t$ lies on $\mathcal{M}$ by design, regardless of the value of A_t.

3.2 Structured Prior over Motion Primitives

We place a prior on A_t that encodes the physical assumption of motion persistence. We model it as Gaussian and condition it on the previous motion:

$$p_\theta(A_t \mid A_{t-1}) = \mathcal{N}\big(A_t;\, \Gamma A_{t-1},\, \Sigma\big), \tag{3}$$

where $\Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_d)$ with $0 \le \gamma_j \le 1$ controls inertia per primitive, and $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_d^2) \succ 0$ is the process covariance. This prior is fixed and does not depend on neural data, so it remains valid under arbitrary degradation of y_t.

3.3 Variational Posterior over Motion

The true motion A_t is latent. We approximate its posterior with a Gaussian distribution whose parameters depend on the current observation and the preceding motion:

$$q_\phi(A_t \mid c_{\le t}, A_{<t}) = \mathcal{N}\big(A_t;\, \mu_{\phi,t},\, \mathrm{diag}(\sigma^2_{\phi,t})\big), \tag{4}$$

where $c_t = (\tilde{x}_t, r_t, A_{t-1})$ denotes the causal context. The diagonal form is essential: since each coordinate of A_t corresponds to a distinct primitive, the posterior variance $\sigma^2_{\phi,t,j}$ quantifies uncertainty about the j-th primitive. Coordinates with small posterior variance are well constrained by evidence; coordinates with large posterior variance are not. This is the mechanism by which primitives can be weighted differently without any hand-specified scheme.

During training, we sample A_t via the reparameterization trick,

$$A_t = \mu_{\phi,t} + \sigma_{\phi,t} \odot \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, I_d), \tag{5}$$

so that gradients propagate through the stochastic motion. At inference, we use $A_t = \mu_{\phi,t}$ and retain $\sigma^2_{\phi,t}$ for uncertainty propagation.

3.4 Adaptive Fusion of Motion and Evidence

The motion proposal $\bar{x}_t$ carries physical structure but may be inaccurate; the observation $\tilde{x}_t$ carries data-driven information but may be unreliable. We fuse them according to their relative uncertainties. The observation covariance is predicted from the reliability features,

$$R_t = \mathrm{diag}\big(\exp(g_\omega(r_t))\big) \in \mathbb{R}^{d\times d}, \tag{6}$$

and the predictive covariance of the motion proposal is

$$P_t^- = P_{t-1} + \Delta t^2\, \mathrm{diag}(\sigma^2_{\phi,t}). \tag{7}$$

Two properties are worth noting. First, P_t^- accumulates posterior variance across primitives, so primitives that are poorly inferred contribute more to predictive uncertainty. Second, its dependence on $\sigma^2_{\phi,t}$ makes the fusion weight adaptive: primitives with high posterior uncertainty are automatically down-weighted.

Assuming conditional independence between the motion proposal and the observation, the fusion weight and updated covariance follow from Gaussian conditioning:

$$W_t = R_t(P_t^- + R_t)^{-1}, \qquad P_t = \big((P_t^-)^{-1} + R_t^{-1}\big)^{-1}. \tag{8}$$

The corrected state is obtained by applying the weighted tangent discrepancy at the observation:

$$\hat{x}_t = \Phi_{\tilde{x}_t}\Big(B_{\tilde{x}_t} W_t B_{\tilde{x}_t}^{-1} \Psi_{\tilde{x}_t}(\bar{x}_t)\Big), \tag{9}$$

where $\Psi_x : \mathcal{M} \to T_x\mathcal{M}$ is the local inverse retraction. The correction is bounded by the geodesic distance between $\tilde{x}_t$ and $\bar{x}_t$, so no single step can produce a physically implausible jump. When R_t is small (reliable evidence), W_t is small and $\hat{x}_t$ stays near $\tilde{x}_t$. When R_t is large (degraded evidence), W_t grows and $\hat{x}_t$ relies on the physical proposal.

3.5 Learning Objective

We train the model by maximizing a sequential evidence lower bound on the behavioral likelihood. The likelihood of the true state given the corrected estimate is

$$p_\theta(x_t \mid A_t, c_{\le t}) \propto \exp\left(-\frac{d^2_{\mathcal{M}}(x_t, \hat{x}_t)}{2\tau^2}\right), \tag{10}$$

where $d_{\mathcal{M}}$ is the geodesic distance and $\tau > 0$ controls the observation scale. The ELBO is

$$\begin{aligned}\mathcal{L}_{\mathrm{ELBO}} = &\sum_{t=1}^{T} \mathbb{E}_{q_\phi}\big[\log p_\theta(x_t \mid A_t, c_{\le t})\big] \\ &- \sum_{t=1}^{T} \mathbb{E}_{q_\phi}\big[\mathrm{KL}\big(q_\phi(A_t \mid c_{\le t}, A_{<t}) \,\|\, p_\theta(A_t \mid A_{t-1})\big)\big].\end{aligned} \tag{11}$$

For diagonal Gaussians, the KL term has the closed form

$$\mathrm{KL}(q_t \| p_t) = \frac{1}{2}\sum_{j=1}^{d}\left[\log\frac{\sigma^2_{p,t,j}}{\sigma^2_{q,t,j}} + \frac{\sigma^2_{q,t,j} + (\mu_{q,t,j} - \mu_{p,t,j})^2}{\sigma^2_{p,t,j}} - 1\right]. \tag{12}$$

The total loss minimized is $\mathcal{L} = -\mathcal{L}_{\mathrm{ELBO}}$.

The likelihood term drives behavioral accuracy and dominates when observations are reliable. The KL term regularizes the posterior toward the structured prior and dominates when observations are weak or missing. The balance between the two is not fixed but emerges from the inference itself, which is precisely what allows the model to remain robust under degradation while preserving accuracy under clean conditions. The overall algorithm is shown in 1.

3.6 Implementation

We instantiate the abstract quantities above as follows. The state space $\mathcal{M}$ is one of $\mathbb{S}^1$, $\mathbb{R}$, or $\mathbb{T}^2$, with the retraction Φ_x, inverse retraction Ψ_x, and geodesic distance $d_{\mathcal{M}}$ chosen accordingly. The local frame B_x is orthonormal. The posterior parameters $\mu_{\phi,t}$ and $\log\sigma^2_{\phi,t}$ are produced by a gated recurrent unit (GRU) followed by a linear head:

$$h_t = \mathrm{GRU}_\phi(h_{t-1}, c_t), \qquad (\mu_{\phi,t}, \log\sigma^2_{\phi,t}) = g_\phi(h_t). \tag{13}$$

The observation covariance mapping g_ω is a multilayer perceptron. The initial state $\hat{x}_0$ and covariance P_0 are set from the first observation. We use one reparameterized sample per step and linearly warm up the KL term during training. At inference, $A_t = \mu_{\phi,t}$ and $\sigma^2_{\phi,t}$ is retained only for uncertainty propagation.

4 Experiment

We evaluate MaPIC across synthetic benchmarks and three real neural datasets. Our experiments address four questions: (i) Do the structured physical primitives and adaptive fusion each contribute to performance? (ii) Does the method improve decoding across different state-space topologies ($\mathbb{S}^1$, $\mathbb{T}^2$, $\mathbb{R}$)? (iii) Does it remain the best method as neurons are dropped? (iv) Is the improvement consistent across diverse base decoders?

4.1 Experimental Setup

Datasets. We use one synthetic benchmark and three real neural datasets. The **synthetic $\mathbb{S}^1$ and $\mathbb{T}^2$** benchmarks provide controlled circular and toroidal state spaces with known ground-truth dynamics, Poisson-tuned neurons, and a tunable analytical-dynamics mismatch κ. Because the synthetic data are generated under fully controlled conditions, we can independently introduce two types of degradation: observation noise and random neuronal dropout. In contrast, real neural recordings already contain inherent observation noise from biological and recording processes; we therefore do not add

Algorithm 1 Training and inference

Initialize $h_0 = A_0 = 0$, and set $\hat{x}_0, P_0$ from the first observation.
for $t = 1, \ldots, T$ **do**
  Compute $\tilde{x}_t = D(F(y_t))$ and $r_t = \rho(y_t)$.
  Form context $c_t = (\tilde{x}_t, r_t, A_{t-1})$.
  Update h_t, then obtain $\mu_{\phi,t}$ and $\log \sigma^2_{\phi,t}$.
  if training **then**
    Sample $A_t \sim q_\phi$.
  else
    Set $A_t = \mu_{\phi,t}$.
  end if
  Compute proposal $\bar{x}_t = \Phi_{\hat{x}_{t-1}}(\Delta t B_{\hat{x}_{t-1}} A_t)$.
  Compute R_t, P_t^-, W_t, P_t.
  Compute corrected state $\hat{x}_t$ via weighted tangent update.
  if training **then**
    Accumulate $\log p_\theta$ and KL.
  end if
end for
if training **then**
  Update θ, ϕ by minimizing $-\mathcal{L}_{\mathrm{ELBO}}$ with KL warm-up.
end if
Return $\hat{x}_{1:T}$.

artificial noise to real data but instead focus on the more realistic challenge of neuronal dropout.The **ADn head-direction** dataset ($\mathbb{S}^1$) contains five sessions of anterior dorsal thalamic nucleus recordings from the CRCNS th-1 dataset Peyrache et al. (2015). The **grid-cell** dataset ($\mathbb{T}^2$) pools 483 cells from modules 1–3 of rat r on day 1 Gardner et al. (2022). The **hippocampal CA1 linear-track** dataset ($\mathbb{R}$) includes four rats (Achilles, Buddy, Cicero, Gatsby) running on a 1.6 m track, following the CEBRA evaluation protocol Grosmark & Buzsáki (2016); Schneider et al. (2023).

Degradation protocols. On synthetic data, we test two axes: (a) **observation noise**, by varying the Poisson noise level and the mismatch κ; and (b) **random neuronal dropout**, by dropping neurons at rates from 0% to 80%. On real data, we test **random neuronal dropout** at rates from 0% to 80%. At each dropout level, we compare MaPIC against all baselines.

Base decoders. We apply MaPIC on top of five base models: CEBRA Schneider et al. (2023), π-VAE Zhou & Wei (2020), DFINE Abbaspourazad et al. (2024), a point-process filter, and a Kalman filter Wu et al. (2003). All base models are trained on the training set and kept frozen; only the temporal correction module is trained.

Metrics. Circular variables: median absolute angular error (radians). Linear and 2D position: median absolute error (cm). Grid cells: median positional error during movement (speed $\geq 2.5\,\mathrm{cm\,s^{-1}}$). Results are averaged over runs with standard deviation.

Baselines for temporal correction. No correction, causal EMA, Kalman filter, vanilla GRU, and ablations of MaPIC (analytic-only, learned-only, no gating, plain bounded correction).

4.2 Controlled Synthetic Benchmarks

We first isolate the two core mechanisms: the structured motion prior and the adaptive fusion. On $\mathbb{S}^1$, the ground-truth velocity is decomposed into an analytical term u_t and an unmodeled residual $\delta_t = \kappa f(\theta_t; \phi)$. On $\mathbb{T}^2$, we repeat the protocol with a capacity-matched Euclidean control.

Results under observation noise. Full MaPIC achieves the lowest geodesic error across all label budgets and mismatch levels. Ablations confirm that both components are necessary: analytic-only is competitive when the analytical dynamics are accurate but degrades rapidly as mismatch grows; learned-only adapts to the mismatch but discards structure and requires more data. Full MaPIC reuses the trustworthy component and learns only the missing residual. On $\mathbb{T}^2$, Torus MaPIC outperforms the Euclidean variant in all 16 combinations of mismatch and label budget. At $(\kappa, B) =$

Table 1: Ablation results for temporal correction methods applied to a frozen CEBRA readout, mean $\pm$ std (cm) over nine runs. Lower is better. Best results in bold.

Method	Achilles	Buddy	Cicero	Gatsby
Official KNN	5.62 ± 0.52	13.66 ± 2.75	21.56 ± 1.49	12.47 ± 2.75
EMA	5.74 ± 0.68	13.72 ± 2.52	21.55 ± 1.51	12.43 ± 2.76
Kalman filter	5.87 ± 0.57	13.78 ± 2.96	22.01 ± 1.39	12.63 ± 2.80
Vanilla GRU	6.32 ± 1.21	13.53 ± 2.90	17.67 ± 3.51	12.85 ± 1.87
MaPIC-Full	$\mathbf{5.37 \pm 0.41}$	$\mathbf{11.70 \pm 2.50}$	$\mathbf{17.93 \pm 3.45}$	$\mathbf{11.30 \pm 2.33}$
MaPIC-NoGate	5.34 ± 0.39	12.45 ± 1.62	21.24 ± 1.76	11.55 ± 2.37
MaPIC-Plain	13.98 ± 13.76	12.75 ± 2.52	17.38 ± 3.45	13.06 ± 1.76

$(1.5, 16)$, branch-cut error is 0.168 rad for Torus MaPIC versus 0.393 rad for Euclidean MaPIC, validating that respecting intrinsic geometry is essential.

Results under neuronal dropout. On synthetic data, we sweep the dropout rate from 0% to 80%. MaPIC achieves the lowest error at every level, with the gap to the best baseline widening monotonically as dropout increases.

4.3 Head-Direction Decoding

We apply MaPIC on top of a frozen CEBRA pointwise decoder on the ADn head-direction dataset. Since real recordings already contain inherent noise, we evaluate robustness through neuronal dropout from 0% to 80%.

MaPIC improves over CEBRA in 14 of 15 paired session-population conditions, with a median relative error reduction of 7.9%. The gain is largest for the sparsest populations and decreases monotonically as more neurons become available: 11.9%, 7.7%, and 5.2% for 5%, 10%, and 30% of neurons, respectively. Across the full dropout range from 0% to 80%, MaPIC achieves the lowest median angular error at every level, consistently outperforming the best baseline. This confirms the intended role of the topological prior: it compensates for ambiguity when instantaneous evidence is limited, and exerts less influence when the neural observation is already informative.

4.4 Grid-Cell Decoding

We evaluate MaPIC on grid-cell recordings, comparing against the raw Bayesian decoder and a planar exponential-moving-average (EMA) baseline. Degradation is again evaluated through neuronal dropout.

MaPIC outperforms both baselines at every training fraction and every dropout level. Relative to the raw Bayesian decoder, it reduces median positional error by 12.3%–16.2%; relative to planar-EMA, by 7.5%–14.9%. Notably, with only 20% of the training data, MaPIC reaches 9.28 cm, outperforming both baselines trained on the full dataset (10.16 cm and 9.40 cm). Under dropout from 0% to 80%, MaPIC achieves the lowest error at every level. This demonstrates that respecting toroidal geometry provides information that planar temporal smoothing cannot recover.

4.5 Position Decoding on a Linear Track

We freeze both the CEBRA representation and its official k-nearest-neighbor decoder, and train only the temporal correction module on hippocampal CA1 recordings. Degradation is evaluated through neuronal dropout.

Neither causal EMA nor Kalman filtering consistently improves the frozen kNN decoder. A vanilla GRU produces a gain for Cicero but increases error or variability for the other animals. In contrast, MaPIC reduces error for all four rats, with relative improvements from 4.4% to 16.8% (average 11.3%). This cross-animal consistency is not reproduced by smoothing or recurrence alone.

Ablations confirm that both adaptive gating and bounded correction are important. Removing either occasionally improves the mean in a single condition but substantially reduces robustness across animals and runs. For example, MaPIC-Plain yields 13.98 ± 13.76 cm on Achilles, indicating high sensitivity across repeated evaluations.

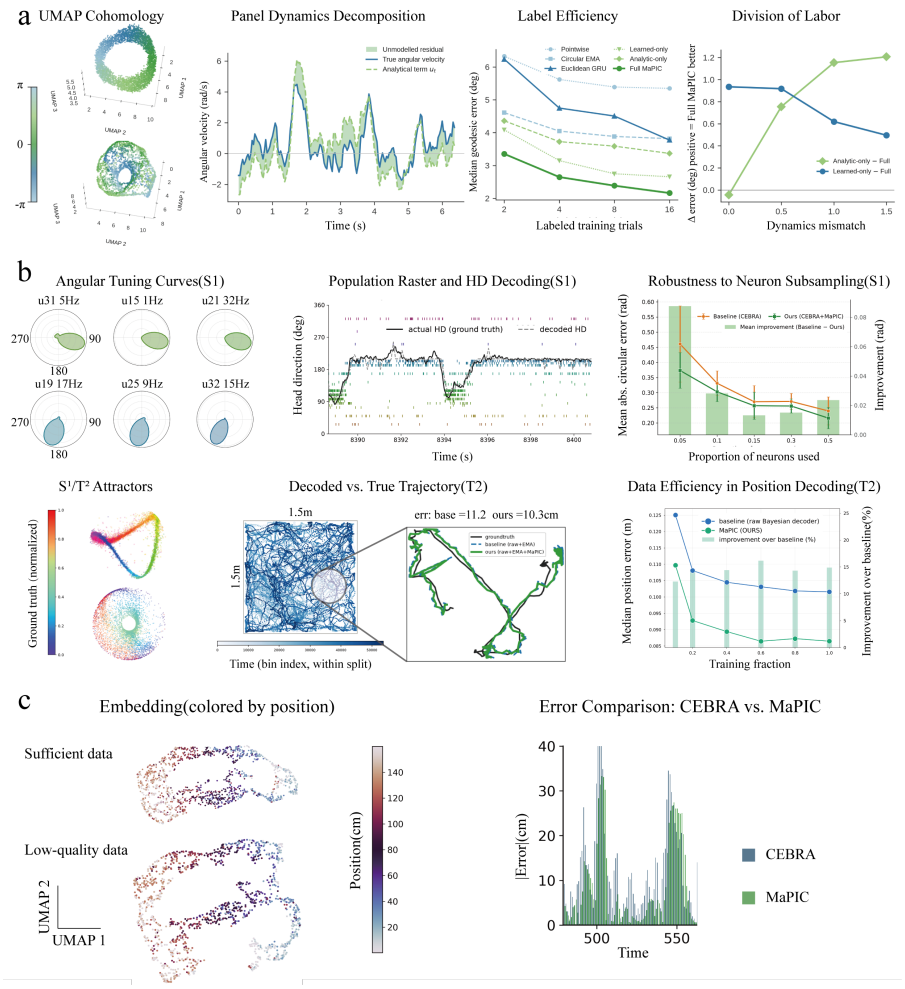


Figure 2: Decoding error as a function of neuronal dropout rate (0% to 80%) on synthetic and real datasets. At every level, MaPIC achieves the lowest error among all methods. The advantage increases with dropout, demonstrating that physical primitives and adaptive fusion provide a stable prior when neural evidence degrades.

4.6 Robustness to Neuronal Dropout

A central claim of this work is that physical primitives provide a stable prior when neural evidence degrades. We randomly drop neurons at test time over different dropout levels. This protocol applies to both synthetic and real datasets.

Across all datasets and all dropout levels from 0% to 80%, MaPIC achieves the lowest median decoding error, consistently outperforming the best baseline at every level. The performance gap widens monotonically as dropout increases: at 0% dropout the improvement is modest, reflecting that neural evidence is already informative; at 80% dropout the improvement becomes substantial, reflecting the growing importance of the structured physical prior and adaptive fusion. MaPIC does not merely degrade gracefully: it provides the best available estimate at every level of neuronal loss, and the benefit of physical primitives is most pronounced exactly when neural evidence is least reliable.

Table 2: Position-decoding error on a 1.6 m linear track, mean ± std (cm) over nine runs. Parentheses show MaPIC's relative improvement. Bold indicates MaPIC-corrected results.

Category	Method	Achilles	Buddy	Cicero	Gatsby
Dynamical	Kalman	12.07 ± 1.26	17.92 ± 2.96	33.59 ± 8.47	22.12 ± 4.38
	Kalman + MaPIC	$\mathbf{11.66 \pm 1.18}$ (3.2% ↓)	$\mathbf{17.01 \pm 2.89}$ (5.1% ↓)	$\mathbf{32.33 \pm 7.17}$ (3.1% ↓)	$\mathbf{20.09 \pm 3.07}$ (7.7% ↓)
	DFINE	5.20 ± 0.79	11.07 ± 2.43	29.26 ± 5.80	13.25 ± 4.07
	DFINE + MaPIC	$\mathbf{5.08 \pm 0.81}$ (2.4% ↓)	$\mathbf{10.55 \pm 1.51}$ (3.2% ↓)	$\mathbf{27.98 \pm 5.98}$ (4.7% ↓)	$\mathbf{13.09 \pm 3.85}$ (0.9% ↓)
Encoder-Decoder	Point-process	9.69 ± 1.32	14.60 ± 1.68	16.24 ± 1.49	14.31 ± 4.01
	Point-process + MaPIC	$\mathbf{9.56 \pm 1.42}$ (1.5% ↓)	$\mathbf{12.49 \pm 2.20}$ (14.5% ↓)	$\mathbf{16.14 \pm 1.55}$ (0.6% ↓)	$\mathbf{13.45 \pm 4.17}$ (6.6% ↓)
	π-VAE	13.10 ± 6.09	28.90 ± 8.71	49.52 ± 11.02	20.75 ± 6.52
	π-VAE + MaPIC	$\mathbf{10.78 \pm 5.32}$ (18.4% ↓)	$\mathbf{25.13 \pm 7.88}$ (13.2% ↓)	$\mathbf{42.01 \pm 11.69}$ (15.3% ↓)	$\mathbf{16.73 \pm 5.07}$ (18.5% ↓)
Contrastive	CEBRA	5.59 ± 0.59	14.00 ± 2.88	20.83 ± 2.11	12.06 ± 2.43
	CEBRA + MaPIC	$\mathbf{5.30 \pm 0.59}$ (5.0% ↓)	$\mathbf{12.71 \pm 2.47}$ (9.0% ↓)	$\mathbf{17.95 \pm 2.66}$ (14.1% ↓)	$\mathbf{11.27 \pm 2.29}$ (6.5% ↓)

4.7 Generalization across Base Models

We evaluate MaPIC with five frozen base models spanning dynamical, encoder-decoder, and contrastive approaches.

As shown in Table 2, MaPIC reduces error in all 5 model-animal combinations. Improvements range from 13.2% to 18.5% for π-VAE, 5.0% to 14.1% for CEBRA, 0.9% to 7.7% for DFINE, 0.6% to 14.5% for point-process, and 3.1% to 7.7% for Kalman. This suggests that MaPIC is not tied to a particular representation-learning objective; it serves as a general temporal readout that adds structured dynamics while preserving the instantaneous evidence supplied by diverse pretrained models.

In summary, MaPIC improves decoding across synthetic benchmarks and three real neural datasets, especially when neural evidence is weak. It remains robust to observation noise and neuronal dropout, achieving the lowest error at every tested dropout level from 0% to 80%. These results support physical motion primitives as stable priors across base decoders and state-space topologies.

5 Discussion

MaPIC combines known motion structure with learned inference to improve neural decoding under limited observations or supervision. Simulations examine the roles of dynamics and geometry, while real-data experiments demonstrate gains with sparse populations, limited training data, and frozen base decoders. Lower mean errors across all 20 base-model–animal combinations support its utility as a reusable temporal readout. These results concern behavioral estimation, not identification of the underlying biological circuitry.

When structured motion helps. MaPIC is suited to tasks with known state geometry, noisy but informative readouts, and useful temporal continuity. Larger gains under sparse head-direction recordings and reduced grid-cell training data support this interpretation. Simulations further suggest that learned correction becomes more valuable as analytical dynamics become less accurate. Benefits may diminish when the base decoder is already accurate.

Scope and practical boundaries. The observed gains depend on informative neural evidence and appropriate motion structure. Prolonged signal loss can leave persistent ambiguity, while inaccurate priors may bias estimates during abrupt changes. Uncertain topology or shifts in the neural observation mapping may require a revised state model or representation adaptation. The current evaluation concerns offline decoding; sustained dropout recovery and closed-loop BCI performance remain to be established.

AI USE STATEMENT

In this work, we used generative AI tools for research ideation and execution, including assisting with the development and refinement of research ideas and providing support for mathematical reasoning and proofs.

We have not used generative AI tools for other tasks with required disclosure, and the rest of the required disclosure tasks are not applicable to this work.

Additionally, we used generative AI tools for retrieval and discovery, including identifying potentially relevant literature and information, and for aiding or polishing the writing of the paper. We have reviewed all AI-assisted work. AI-assisted research ideas and mathematical arguments were independently checked and verified by the authors, references and factual information suggested by AI tools were manually verified against the original sources, and AI-assisted writing was reviewed and edited by the authors for correctness and clarity. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Hamidreza Abbaspourazad, Eray Erturk, Bijan Pesaran, and Maryam M. Shanechi. Dynamical flexible inference of nonlinear latent factors and structures in neural population activity. *Nature Biomedical Engineering*, 8(1):85–108, 2024. doi: 10.1038/s41551-023-01106-1.

Mehdi Aghagolzadeh and Wilson Truccolo. Inference and decoding of motor cortex low-dimensional dynamics via latent state-space models. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(2):272–282, 2016. doi: 10.1109/TNSRE.2015.2470527.

David M Brandman, Tommy Hosman, Jad Saab, Michael C Burkhart, Benjamin E Shanahan, John G Ciancibello, Anish A Sarma, Daniel J Milstein, Carlos E Vargas-Irwin, Brian Franco, Jessica Kelemen, Christine Blabe, Brian A Murphy, Daniel R Young, Francis R Willett, Chethan Pandarinath, Sergey D Stavisky, Robert F Kirsch, Benjamin L Walter, A Bolu Ajiboye, Sydney S Cash, Emad N Eskandar, Jonathan P Miller, Jennifer A Sweet, Krishna V Shenoy, Jaimie M Henderson, Beata Jarosiewicz, Matthew T Harrison, John D Simeral, and Leigh R Hochberg. Rapid calibration of an intracortical brain–computer interface for people with tetraplegia. *Journal of Neural Engineering*, 15(2):026007, 2018. doi: 10.1088/1741-2552/aa9ee7.

Michael C Burkhart, David M Brandman, Brian Franco, Leigh R Hochberg, and Matthew T Harrison. The discriminative Kalman filter for Bayesian filtering with nonlinear and nongaussian observation models. *Neural Computation*, 32(5):969–1017, 2020. doi: 10.1162/neco_a_01275.

Rishidev Chaudhuri, Berk Gerçek, Biraj Pandey, Adrien Peyrache, and Ila Fiete. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature neuroscience*, 22(9):1512–1520, 2019.

Mark M Churchland, John P Cunningham, Matthew T Kaufman, Justin D Foster, Paul Nuyujukian, Stephen I Ryu, and Krishna V Shenoy. Neural population dynamics during reaching. *Nature*, 487 (7405):51–56, 2012. doi: 10.1038/nature11129.

John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11):1500–1509, 2014. doi: 10.1038/nn.3776.

Juan A Gallego, Matthew G Perich, Lee E Miller, and Sara A Solla. Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nature Communications*, 9:4233, 2018. doi: 10.1038/s41467-018-06560-9.

Richard J Gardner, Erik Hermansen, Marius Pachitariu, Yoram Burak, Nils A Baas, Benjamin A Dunn, May-Britt Moser, and Edvard I Moser. Toroidal topology of population activity in grid cells. *Nature*, 602(7895):123–128, 2022.

Sigurd Gaukstad, Melvin Vaupel, Valdemar Kargård Olsen, Erik Hermansen, and Benjamin Adric Dunn. Learning coherent representations: A topological approach to interpretability. In *Forty-third International Conference on Machine Learning*, 2026. URL `https://openreview.net/forum?id=VnlpOuA3Bs`.

Vikash Gilja, Paul Nuyujukian, Cynthia A Chestek, John P Cunningham, Byron M Yu, Joline M Fan, Mark M Churchland, Matthew T Kaufman, Jonathan C Kao, Stephen I Ryu, and Krishna V Shenoy. A high-performance neural prosthesis enabled by control algorithm design. *Nature Neuroscience*, 15(12):1752–1757, 2012. doi: 10.1038/nn.3265.

Adam Gosztolai, Robert L. Peach, Alexis Arnaudon, Mauricio Barahona, and Pierre Vandergheynst. MARBLE: interpretable representations of neural population dynamics using geometric deep learning. *Nature Methods*, 22(3):612–620, 2025. doi: 10.1038/s41592-024-02582-2.

Andres D Grosmark and György Buzsáki. Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. *Science*, 351(6280):1440–1443, 2016.

Han-Lin Hsieh and Maryam Shanechi. Probabilistic geometric principal component analysis with application to neural data. In *International Conference on Learning Representations*, volume 2025, pp. 54124–54142, 2025.

Peter J Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35 (1):73–101, 1964. doi: 10.1214/aoms/1177703732.

Cole Hurwitz, Akash Srivastava, Kai Xu, Justin Jude, Matthew Perich, Lee Miller, and Matthias H Hennig. Targeted neural dynamical modeling. In *Advances in Neural Information Processing Systems*, volume 34, pp. 29379–29392, 2021.

Kristopher Jensen, Ta-Chu Kao, Marco Tripodi, and Guillaume Hennequin. Manifold gplvms for discovering non-euclidean latent structure in neural data. *Advances in Neural Information Processing Systems*, 33:22580–22592, 2020.

Maximilian Karl, Maximilian Soelch, Justin Bayer, and Patrick van der Smagt. Deep variational Bayes filters: Unsupervised learning of state space models from raw data. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=HyTqHL5xg.

Mikail Khona and Ila R Fiete. Attractor and integrator networks in the brain. *Nature Reviews Neuroscience*, 23(12):744–766, 2022.

Rahul G Krishnan, Uri Shalit, and David Sontag. Deep Kalman filters. *arXiv preprint arXiv:1511.05121*, 2015. doi: 10.48550/arXiv.1511.05121.

C J Masreliez and R D Martin. Robust Bayesian estimation for the linear model and robustifying the Kalman filter. *IEEE Transactions on Automatic Control*, 22(3):361–371, 1977. doi: 10.1109/TAC.1977.1101538.

Chethan Pandarinath, Daniel J O'Shea, Jasmine Collins, Rafal Jozefowicz, Sergey D Stavisky, Jonathan C Kao, Eric M Trautmann, Matthew T Kaufman, Stephen I Ryu, Leigh R Hochberg, et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature methods*, 15(10):805–815, 2018.

Adrien Peyrache, Marie M Lacroix, Peter C Petersen, and György Buzsáki. Internally organized mechanisms of the head direction sense. *Nature neuroscience*, 18(4):569–575, 2015.

Omid G Sani, Hamidreza Abbaspourazad, Yan T Wong, Bijan Pesaran, and Maryam M Shanechi. Modeling behaviorally relevant neural dynamics enabled by preferential subspace identification. *Nature Neuroscience*, 24(1):140–149, 2021. doi: 10.1038/s41593-020-00733-0.

Omid G Sani, Bijan Pesaran, and Maryam M Shanechi. Dissociative and prioritized modeling of behaviorally relevant neural dynamics using recurrent neural networks. *Nature neuroscience*, 27 (10):2033–2045, 2024.

Steffen Schneider, Jin Hwa Lee, and Mackenzie Weygandt Mathis. Learnable latent embeddings for joint behavioural and neural analysis. *Nature*, May 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06031-6. URL https://doi.org/10.1038/s41586-023-06031-6.

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.

Wei Wu, Michael J Black, Yun Gao, Elie Bienenstock, Mijail Serruya, Ammar Shaikhouni, and John P Donoghue. Neural decoding of cursor motion using a Kalman filter. In *Advances in Neural Information Processing Systems*, volume 15, pp. 133–140. MIT Press, 2003.

Weiheng Zhong and Hadi Meidani. Pi-vae: Physics-informed variational auto-encoder for stochastic differential equations. *Computer Methods in Applied Mechanics and Engineering*, 403:115664, 2023. ISSN 0045-7825. doi: https://doi.org/10.1016/j.cma.2022.115664. URL https://www.sciencedirect.com/science/article/pii/S0045782522006193.

Ding Zhou and Xue-Xin Wei. Learning identifiable and interpretable latent models of high-dimensional neural activity using pi-vae. *Advances in neural information processing systems*, 33:7234–7247, 2020.

A Appendix

A.1 Synthetic $\mathbb{S}^1$ data generation and experimental protocol

This section specifies the synthetic $\mathbb{S}^1$ head-direction trajectories, neural population activity, controlled analytical-dynamics mismatch, and evaluation protocol used in the main experiments. The design separates four components: a frozen instantaneous neural readout, an available analytical self-motion term, a tangent-space residual that must be learned, and a bounded manifold correction anchored at the current neural evidence.

A.1.1 Circular trajectories and ground-truth dynamics

The decoding target is the head direction

$$\theta_t \in \mathbb{S}^1 \simeq [-\pi, \pi). \tag{14}$$

All angular operations use the principal-value map

$$\mathrm{wrap}(x) = ((x + \pi) \bmod 2\pi) - \pi \in [-\pi, \pi). \tag{15}$$

Each trial contains $T = 160$ time points with step size $\Delta t = 0.04\,\mathrm{s}$, corresponding to a duration of $T\Delta t = 6.4\,\mathrm{s}$. Initial angles are sampled independently from $\mathcal{U}(-\pi, \pi)$, and initial angular velocities are sampled from $\mathcal{N}(0, 1)$ in radians per second.

The default controlled generator uses a smooth angular-velocity process with occasional rapid turns. Define the autoregressive coefficient $\alpha_\omega = \exp(-\Delta t/0.45)$. The dynamics are

$$\begin{aligned} \omega_t &= \alpha_\omega \omega_{t-1} + 1.35\sqrt{1-\alpha_\omega^2}\,\xi_t + J_t, \\ \theta_t &= \mathrm{wrap}(\theta_{t-1} + \Delta t\,\omega_t + 0.025\,\varepsilon_t), \end{aligned} \tag{16}$$

where $\xi_t, \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. The jump term J_t introduces an abrupt turn independently at each time step with probability 0.015. Conditional on a turn, its sign is sampled uniformly from $\{-1, +1\}$ and its magnitude from $\mathcal{U}(1.5, 3.0)$; otherwise, $J_t = 0$. This process produces predominantly smooth circular motion with a small number of rapid directional changes.

A.1.2 Controlled analytical-dynamics mismatch

To distinguish the motion component available to the analytical model from an unmodeled residual, we define an angle-dependent tangent velocity field

$$f(\theta; \varphi) = 0.90\sin(2\theta + \varphi) + 0.35\sin(\theta - 0.5\varphi) + 0.20\cos(3\theta + 0.25), \tag{17}$$

where the phase $\varphi \sim \mathcal{U}(-\pi, \pi)$ is sampled once and held fixed within each random seed. For mismatch level $\kappa \in \{0, 0.5, 1.0, 1.5\}$, the ground-truth unmodeled velocity residual and the analytical term supplied to the model are

$$\delta_t = \kappa f(\theta_t; \varphi), \qquad u_t = \omega_t - \delta_t. \tag{18}$$

Hence, the deterministic angular velocity admits the exact decomposition $\omega_t = u_t + \delta_t$. The process noise added directly to the angular update in Equation equation 16 belongs to neither u_t nor δ_t. At $\kappa = 0$, the supplied analytical term equals the deterministic angular velocity. Increasing κ progressively strengthens the systematic mismatch between u_t and the ground-truth dynamics.

The comparison across mismatch levels is strictly paired. For each random seed, we first generate a base dataset and then reuse exactly the same θ_t trajectories and neural observations at every value of κ; only u_t and the corresponding residual values in Equation equation 18 change. Performance differences across mismatch levels therefore cannot be attributed to changes in test trajectories, neural noise, or neuronal tuning. The ground-truth residual is used only for diagnostic statistics. Its supervision weight is set to zero in all reported experiments, and δ_t is never provided to the model as a training target.

A.1.3 Neural population activity

Each trajectory is encoded by $n_y = 64$ circularly tuned neurons. The nominal preferred directions are uniformly spaced on $[-\pi, \pi)$ and independently perturbed by Gaussian noise:

$$\mu_i = \mathrm{wrap}\left(-\pi + \frac{2\pi(i-1)}{n_y} + \nu_i\right), \qquad \nu_i \sim \mathcal{N}(0, 0.05^2), \qquad i \in \{1, \dots, n_y\}. \tag{19}$$

The gain of neuron i is

$$A_i = 2.5 \exp(0.15\zeta_i), \qquad \zeta_i \sim \mathcal{N}(0, 1). \tag{20}$$

Given angle θ_t, its conditional mean count is

$$\lambda_{t,i} = 0.15 + A_i \exp\left[2.0\big(\cos(\theta_t - \mu_i) - 1\big)\right], \tag{21}$$

and the observed neural counts are sampled independently as

$$y_{t,i} \sim \mathrm{Poisson}(\lambda_{t,i}). \tag{22}$$

The preferred directions, gains, and mismatch-field phase remain fixed within the dataset generated by a given random seed. This construction yields population activity that varies continuously over $\mathbb{S}^1$ while introducing instantaneous observation noise through finite population size and Poisson sampling.

A.1.4 Data splits and label budgets

For each random seed, we generate 40 independent trials and assign them in generation order to 16 candidate training trials, 8 validation trials, and 16 test trials. We evaluate the label budgets

$$B \in \{2, 4, 8, 16\}, \tag{23}$$

where budget B uses only the first B trials from the candidate training set to fit the instantaneous readout and all additional learnable components. The training sets are therefore nested across budgets, whereas the validation and test sets remain fixed across all budgets and mismatch levels. Validation labels are used only for hyperparameter selection; the test set is never used for model fitting or selection.

The frozen instantaneous decoder is first fitted by ridge regression from the n_y-dimensional neural counts to the circular embedding $(\cos\theta_t, \sin\theta_t)$, using a fixed regularization coefficient of 10:

$$(\widehat{c}_t, \widehat{s}_t) = D_{\mathrm{ridge}}(y_t) \approx (\cos\theta_t, \sin\theta_t), \qquad \widetilde{\theta}_t = \mathrm{atan2}(\widehat{s}_t, \widehat{c}_t), \qquad r_t = \sqrt{\widehat{c}_t^{\,2} + \widehat{s}_t^{\,2}}. \tag{24}$$

The vector magnitude r_t is used as a reliability feature for the current observation; it is not assumed to be a calibrated probability. After fitting on the training set associated with a given label budget, the instantaneous decoder remains frozen.

A.1.5 Training and evaluation protocol

All methods use the same data splits and frozen instantaneous decoder. MaPIC uses a GRU with 32 hidden units and a bounded correction radius of $\rho = 35^\circ$. We train for 80 epochs using AdamW with learning rate 2×10^{-3}, weight decay 10^{-4}, and gradient-norm clipping at 1. The global correction

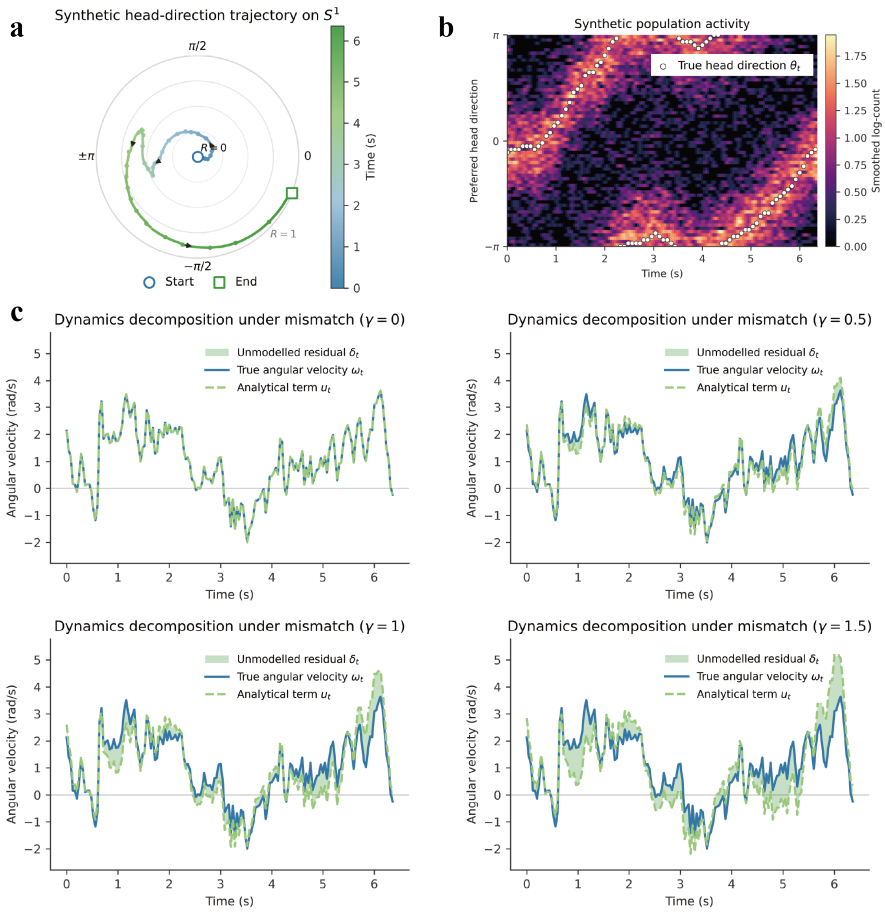


Figure 3: **Synthetic $\mathbb{S}^1$ data and controlled dynamics mismatch.** **(a)** A synthetic head-direction trajectory. Polar angle denotes the ground-truth head direction θ_t, and radius $R = t/T$ denotes normalized time. **(b)** Neural population activity sorted by preferred direction. Color represents Poisson counts smoothed and log-transformed for visualization only; white markers indicate the ground-truth head direction θ_t. **(c)** Decomposition of the dynamics at different mismatch levels. The systematic discrepancy between the supplied analytical dynamics and the ground-truth dynamics increases with κ.

strength is selected from $\beta \in \{0, 0.1, 0.25, 0.5, 0.75, 1.0\}$ using only the median circular error on the validation set. The first frame of each trial is initialized by the instantaneous decoder, and recurrent states are reset at trial boundaries.

The primary metric is the median circular geodesic error over all time points and trials in the test set,

$$E_{\text{test}} = \underset{(j,t)\in\mathcal{D}_{\text{test}}}{\text{median}} \left|\text{wrap}\left(\widehat{\theta}_t^{(j)} - \theta_t^{(j)}\right)\right|. \tag{25}$$

We report E_{test} in degrees. Experiments cover every combination of label budget $B \in \{2, 4, 8, 16\}$ and mismatch level $\kappa \in \{0, 0.5, 1.0, 1.5\}$. All methods share exactly the same training, validation, and test samples within each experimental condition.

A.2 Synthetic $\mathbb{T}^2$ data: Does MaPIC operate intrinsically on the manifold?

We complement the main experiments with a synthetic $\mathbb{T}^2$ study that tests a more specific question than whether MaPIC improves decoding: does its recurrent predict–correct procedure follow the intrinsic geometry of the torus, or does it merely perform Euclidean temporal smoothing on two angular coordinates? Following the design of the $\mathbb{S}^1$ experiment, we hold the instantaneous neural readout, data splits, and learnable architecture fixed and vary only the geometry used by the prediction and correction steps. Differences between Torus MaPIC and Euclidean MaPIC can therefore be attributed to the logarithm map, tangent-space correction, and exponential map on $\mathbb{T}^2$ rather than to model capacity or additional supervision.

Notation. The ground-truth torus state is $\boldsymbol{\theta}_t \in \mathbb{T}^2$; the frozen pointwise readout is $\widetilde{\boldsymbol{\theta}}_t$, the dynamical proposal is $\bar{\boldsymbol{\theta}}_t$, and the final evidence-anchored estimate is $\widehat{\boldsymbol{\theta}}_t$. We write $\boldsymbol{v}_t$ for the ground-truth tangent velocity, $\boldsymbol{u}_t$ for the analytical velocity available to the model, and $\boldsymbol{\delta}_t$ for the unmodeled residual, so that $\boldsymbol{v}_t = \boldsymbol{u}_t + \boldsymbol{\delta}_t$. The scalar κ controls dynamics mismatch, B denotes the number of labeled training trials, and ρ is the maximum correction radius. All vector-valued trigonometric and wrapping operations below are applied componentwise unless stated otherwise.

A.2.1 $\mathbb{T}^2$ trajectories with coupled dynamics

The decoding target is the product of two periodic phases,

$$\boldsymbol{\theta}_t = (\theta_{1,t}, \theta_{2,t}) \in \mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1 \simeq [-\pi, \pi)^2. \tag{26}$$

Each phase is mapped to its principal interval by

$$\text{wrap}(x) = ((x + \pi) \bmod 2\pi) - \pi \in [-\pi, \pi). \tag{27}$$

Each trial contains $T = 200$ time points with $\Delta t = 0.04\,\text{s}$, corresponding to a duration of $T\Delta t = 8\,\text{s}$. Initial phases are sampled from the uniform distribution on $[-\pi, \pi)^2$, and initial velocities are sampled from $\mathcal{N}(\mathbf{0}, I_2)$.

Let $\alpha_v = \exp(-\Delta t/0.50)$. The velocity combines a smooth stochastic process with weak cross-axis coupling:

$$\begin{aligned}
\boldsymbol{v}_t^{(0)} &= \alpha_v \boldsymbol{v}_{t-1} + 1.45\sqrt{1-\alpha_v^2}\,\boldsymbol{\xi}_t, \\
v_{1,t} &= v_{1,t}^{(0)} + 0.20 \sin\theta_{2,t-1}, \qquad v_{2,t} = v_{2,t}^{(0)} + 0.18\cos\theta_{1,t-1}, \\
\boldsymbol{\theta}_t &= \text{wrap}(\boldsymbol{\theta}_{t-1} + \Delta t\, \boldsymbol{v}_t + 0.018\boldsymbol{\varepsilon}_t)\,,
\end{aligned} \tag{28}$$

where $\boldsymbol{\xi}_t, \boldsymbol{\varepsilon}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, I_2)$. Each velocity component is clipped to $[-6, 6]\ \text{rad}\,\text{s}^{-1}$. The cross-axis terms couple motion in the two phase directions, preventing the generator from reducing to two independent $\mathbb{S}^1$ simulations.

A.2.2 Paired construction of analytical-dynamics mismatch

As in the $\mathbb{S}^1$ experiment, we decompose the deterministic tangent velocity exactly into an analytical term supplied to the model and an unmodeled residual:

$$\boldsymbol{v}_t = \boldsymbol{u}_t + \boldsymbol{\delta}_t, \qquad \boldsymbol{\delta}_t = \kappa \boldsymbol{f}(\boldsymbol{\theta}_t; \boldsymbol{\phi}), \tag{29}$$

where $\phi = (p, q)$ is sampled once and held fixed within each random seed, and

$$\boldsymbol{f}(\boldsymbol{\theta}; \phi) = \begin{bmatrix} 0.85\sin(2\theta_1 + p) + 0.35\sin(\theta_1 - \theta_2 + q) \\ 0.75\cos(2\theta_2 - q) + 0.40\sin(\theta_1 + \theta_2 + p) \end{bmatrix}. \tag{30}$$

We evaluate $\kappa \in \{0, 0.5, 1.0, 1.5\}$ and provide the model only with $\boldsymbol{u}_t = \boldsymbol{v}_t - \boldsymbol{\delta}_t$. The ground-truth residual is used only for visualization and diagnostics, never as a supervised training target.

The mismatch comparison is strictly paired. Within each random seed, all values of κ reuse exactly the same ground-truth trajectories and neural observations; only the analytical input $\boldsymbol{u}_t$ changes. Performance differences across mismatch levels therefore cannot be explained by variation in test trajectories or neural noise.

A.2.3 Neural population with mixed toroidal tuning

Each trajectory is encoded by $n_y = 96$ neurons. For neuron i, the preferred phase $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2})$ is sampled uniformly from $[-\pi, \pi)^2$, its gain is $A_i = 2.8\exp(0.18\zeta_i)$ with $\zeta_i \sim \mathcal{N}(0, 1)$, and its mixed-tuning coefficient is sampled as $c_i \sim \mathcal{U}(0.35, 0.75)$. Given $\boldsymbol{\theta}_t$, the conditional mean count is

$$\lambda_{t,i} = 0.15 + A_i \exp\Big\{1.5\big[\cos(\theta_{1,t} - \mu_{i1}) + \cos(\theta_{2,t} - \mu_{i2}) - 2 + c_i\big(\cos[(\theta_{1,t} - \mu_{i1}) + (\theta_{2,t} - \mu_{i2})] - 1\big)\big]\Big\}, \tag{31}$$

and observations are sampled independently as $y_{t,i} \sim \mathrm{Poisson}(\lambda_{t,i})$. The joint-phase term makes individual neurons depend on both circular coordinates; the population thus encodes a two-dimensional torus rather than two disjoint sets of circularly tuned neurons.

Figure 4**(a)** provides a visualization independent of decoding error. A three-dimensional UMAP embedding of the population activity forms a continuous low-dimensional periodic structure, and coloring the same embedding by θ_1 and θ_2 produces continuous gradients along distinct circular directions, consistent with the intended $\mathbb{T}^2$ organization.

A.2.4 Frozen instantaneous readout and controls

For each random seed, we generate 40 trials and split them in fixed generation order into 16 candidate training trials, 8 validation trials, and 16 test trials. We evaluate label budgets $B \in \{2, 4, 8, 16\}$. At budget B, only the first B candidate training trials are used, yielding nested training sets across budgets.

The frozen instantaneous ridge decoder maps neural counts to

$$(\widehat{c}_{1,t}, \widehat{s}_{1,t}, \widehat{c}_{2,t}, \widehat{s}_{2,t}) \approx (\cos\theta_{1,t}, \sin\theta_{1,t}, \cos\theta_{2,t}, \sin\theta_{2,t}), \tag{32}$$

after which $\mathrm{atan2}$ produces the pointwise estimate $\widetilde{\boldsymbol{\theta}}_t \in \mathbb{T}^2$. The ridge coefficient is selected from $\{0.1, 1, 10, 100\}$ using only the validation set. The two cosine–sine magnitudes,

$$\boldsymbol{r}_t = \left(\sqrt{\widehat{c}_{1,t}^2 + \widehat{s}_{1,t}^2}, \sqrt{\widehat{c}_{2,t}^2 + \widehat{s}_{2,t}^2}\right), \tag{33}$$

are used as axis-wise reliability features. They are not assumed to be calibrated probabilities. Once fitted, the instantaneous decoder remains frozen for all temporal methods.

We compare four methods: (i) the frozen frame-wise ridge decoder; (ii) Circular EMA, which independently averages unit-vector representations of the two circular coordinates with a coefficient selected on the validation set; (iii) Torus MaPIC; and (iv) Euclidean MaPIC. The two MaPIC variants have identical inputs, GRUs, hidden dimension (48), residual, inertia, and gating heads, correction radius $\rho = 0.65\,\mathrm{rad}$, optimizer, and training duration. They differ only in the geometry used to compute state differences and retractions.

Under the product metric, the logarithm and exponential maps on $\mathbb{T}^2$ act componentwise:

$$\begin{aligned} \mathrm{Log}_{\boldsymbol{x}}(\boldsymbol{z}) &= \mathrm{atan2}(\sin(\boldsymbol{z} - \boldsymbol{x}), \cos(\boldsymbol{z} - \boldsymbol{x}))\,, \\ \mathrm{Exp}_{\boldsymbol{x}}(\boldsymbol{\eta}) &= \mathrm{wrap}(\boldsymbol{x} + \boldsymbol{\eta}). \end{aligned} \tag{34}$$

For a tangent vector $\boldsymbol{e} \in \mathbb{R}^2$, we use the smooth radial saturation

$$\mathrm{Sat}_\rho(\boldsymbol{e}) = \begin{cases} \rho \tanh(\|\boldsymbol{e}\|_2/\rho)\dfrac{\boldsymbol{e}}{\|\boldsymbol{e}\|_2}, & \boldsymbol{e} \neq \mathbf{0}, \\ \mathbf{0}, & \boldsymbol{e} = \mathbf{0}. \end{cases} \tag{35}$$

Torus MaPIC first combines the analytical velocity and learned residual in $T_{\widehat{\boldsymbol{\theta}}_{t-1}}\mathbb{T}^2$ to form a dynamical proposal $\bar{\boldsymbol{\theta}}_t$. It then evaluates $\mathrm{Log}_{\widetilde{\boldsymbol{\theta}}_t}(\bar{\boldsymbol{\theta}}_t)$, the shortest periodic displacement from current neural evidence to the proposal. After smooth radial saturation and reliability-dependent gating, the correction is mapped back to the torus by $\mathrm{Exp}_{\widetilde{\boldsymbol{\theta}}_t}$. Thus, every proposal, correction direction, and final estimate is defined on $\mathbb{T}^2$ or its tangent spaces, and $\widehat{\boldsymbol{\theta}}_t \in \mathbb{T}^2$ by construction.

Euclidean MaPIC replaces Equation equation 34 with the coordinate difference $\boldsymbol{z}-\boldsymbol{x}$ and unwrapped addition $\boldsymbol{x}+\boldsymbol{\eta}$; all other components remain unchanged. This capacity-matched negative control is crucial. For example, $-\pi+\epsilon$ and $\pi-\epsilon$ are separated by approximately 2ϵ on the torus but by approximately $2\pi-2\epsilon$ in the chosen Euclidean chart. If MaPIC benefits from intrinsic manifold operations, the advantage of Torus MaPIC should be largest near these branch cuts and cannot be explained solely by recurrence or temporal smoothing.

A.2.5 Training, metrics, and results

We train MaPIC for 80 epochs using AdamW with learning rate 2×10^{-3} and weight decay 10^{-4}. The global correction strength is selected from $\beta \in \{0, 0.1, 0.25, 0.5, 0.75, 1.0\}$ using only the validation set. The first frame of each trial is initialized by the frozen instantaneous decoder, and recurrent states are reset at trial boundaries.

The primary metric is the median intrinsic torus error over all test trials and time points,

$$E_{\mathbb{T}^2} = \underset{(j,t)\in\mathcal{D}_{\mathrm{test}}}{\mathrm{median}} \left\| \mathrm{wrap}\left(\widehat{\boldsymbol{\theta}}_t^{(j)} - \boldsymbol{\theta}_t^{(j)}\right) \right\|_2 . \tag{36}$$

We additionally report the error in intrinsic increments between consecutive frames. Frames for which $|\theta_{k,t}| > 0.85\pi$ for at least one $k \in \{1, 2\}$ define the branch-cut neighborhood, on which we separately compute the median intrinsic error. All reported results use the 16 test trials excluded from fitting and model selection, corresponding to 3,200 frames per condition. We average the per-seed metrics over three random seeds.

As shown in Figure 4**(d)**, Torus MaPIC achieves the lowest mean median intrinsic error in all 16 combinations of four mismatch levels and four label budgets. Under the strongest mismatch, $\kappa = 1.5$, increasing the budget from $B = 2$ to $B = 16$ reduces its error from 0.332 to $0.173\,\mathrm{rad}$. At $B = 16$, the corresponding errors of the frozen ridge decoder, Circular EMA, and Euclidean MaPIC are 0.317, 0.225, and $0.221\,\mathrm{rad}$, respectively.

The branch-cut analysis is more diagnostic. At $(\kappa, B) = (1.5, 16)$, Torus MaPIC attains a branch-cut error of $0.168\,\mathrm{rad}$, compared with $0.393\,\mathrm{rad}$ for Euclidean MaPIC. Even with matched analytical dynamics, at $(\kappa, B) = (0, 16)$, the corresponding errors are 0.148 and $0.470\,\mathrm{rad}$. This gap rules out the explanation that an otherwise identical GRU merely learns to smooth two Euclidean angle coordinates.

The intrinsic increment error yields the same conclusion. At $(\kappa, B) = (1.5, 16)$, the median increment error of Torus MaPIC is $0.063\,\mathrm{rad}$, compared with $0.147\,\mathrm{rad}$ for Euclidean MaPIC, $0.228\,\mathrm{rad}$ for Circular EMA, and $0.444\,\mathrm{rad}$ for the frozen ridge decoder. The experiment therefore provides a direct mechanism check, not only an accuracy comparison: MaPIC constructs its proposal on $\mathbb{T}^2$, computes the shortest periodic correction in the tangent space of current neural evidence, and uses the exponential map to preserve a valid torus-valued output. In this controlled setting, correct toroidal geometry is necessary for robust updates across coordinate branch cuts.

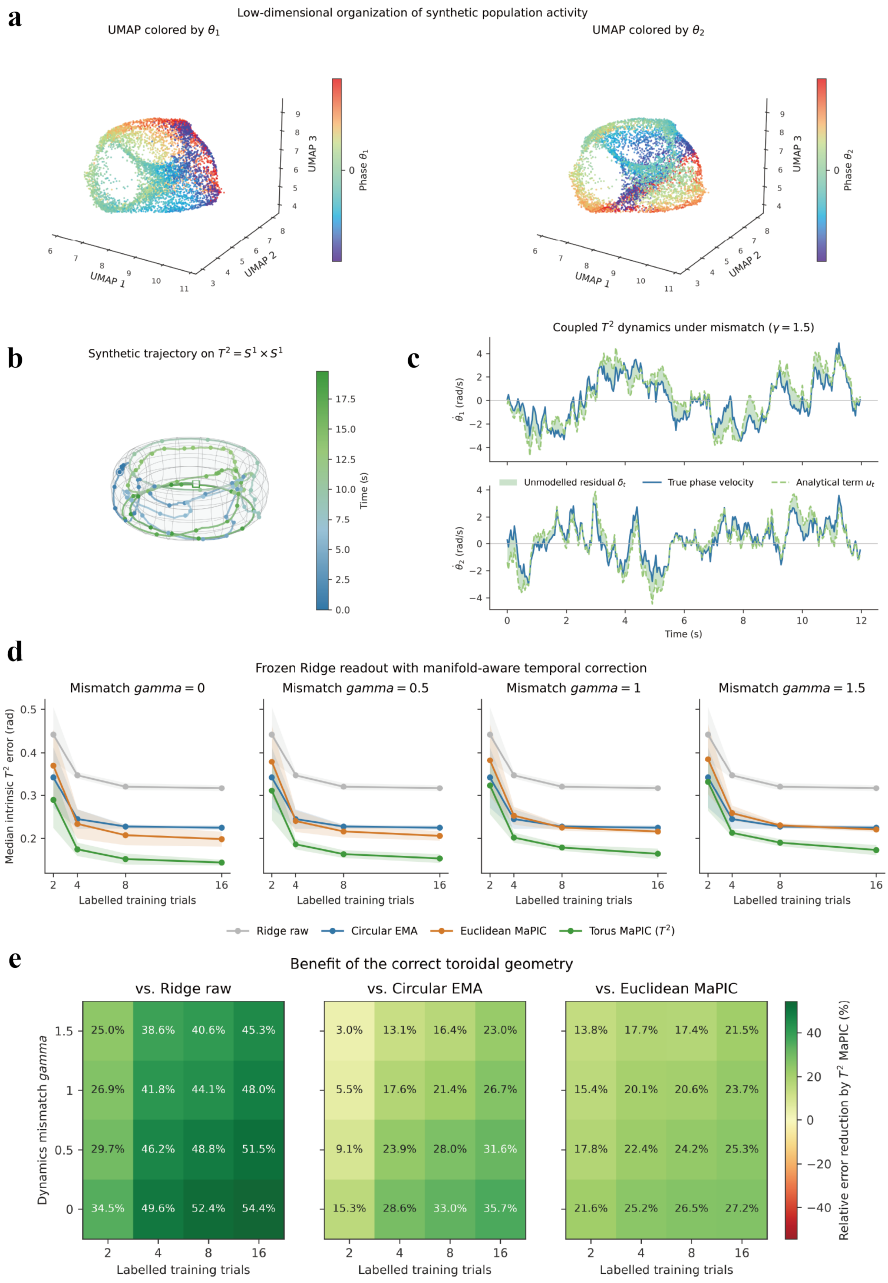


Figure 4: **Mechanism test of MaPIC's manifold operations on synthetic $\mathbb{T}^2$ data. (a)** Three-dimensional UMAP embeddings of synthetic neural population activity, colored separately by the two ground-truth phases θ_1 and θ_2. **(b)** One representative $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ trajectory, colored by time. **(c)** Ground-truth velocity, analytical input supplied to the model, and unmodeled residual along both phase axes under strong dynamics mismatch ($\kappa = 1.5$). **(d)** Intrinsic test error on $\mathbb{T}^2$ across mismatch levels and label budgets. Shading denotes the standard deviation across three random seeds. **(e)** Error reduction achieved by Torus MaPIC relative to the frozen ridge decoder, Circular EMA, and capacity-matched Euclidean MaPIC. Torus MaPIC performs best in every condition.