PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Article submitted to journal

Subject Areas:
computational neuroscience, systems neuroscience, neural dynamics, learning, memory

Keywords:
hippocampus, RNN, CA3, sequence generation, replay, prediction, symmetry breaking, inductive bias

Author for correspondence:
Margot Wagner
e-mail: mwagner@ucsd.edu

Balancing Stability and Flow in Hippocampal Networks via Inductive Bias and Learned Symmetry Breaking

Margot Wagner[1], Yusi Chen[2], Arjun Karuvally[3], Mia Cameron[4], and Terrence J. Sejnowski[1,3]

[1]Institute for Neural Computation, University of California San Diego, La Jolla, CA
[2]Institute of Neuroscience, Chinese Academy of Sciences, Shanghai, China
[3]Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA
[4]Computational Neuroscience Center, University of Washington, Seattle, WA

The hippocampus must balance stable memory representations with internally generated sequential dynamics underlying replay and prediction. How hippocampal circuitry achieves both remains unclear. Here, we show that recurrent neural networks trained on prediction tasks converge to a mixed-symmetry dynamical regime, in which dominant symmetric recurrence stabilizes an attractor while a weaker antisymmetric component induces directed flow. This structure accelerates learning and supports robust replay and prediction. We further show that such symmetry breaking can arise from biologically plausible spike-timing-dependent plasticity (STDP) rules, yielding a tilted Mexican-hat connectivity profile without explicit architectural constraints. Initializing networks with CA3-like structured connectivity biases learning toward this regime, and improves optimization efficiency and performance. These results suggest that hippocampal computation reflects an interaction between biologically constrained circuit structure and task-driven learning, with partial symmetry breaking providing a low-dimensional control principle for balancing stability and flow in sequence generation.

© The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

THE ROYAL SOCIETY PUBLISHING

1. Introduction

(a) Hippocampal memory: stable attractors and directed sequential activation

The hippocampus plays a central role in learning, storing, and retrieving sequential experience. From early spatial map theories [1] to contemporary views of the hippocampus as a generator of predictive internal models [2,3], a central challenge has been to reconcile memory stability with temporally ordered neural activity. In particular, how can neural circuits preserve memories while also supporting structured sequences over time? The hippocampus and especially its sub-region CA3 have long been proposed as recurrent associative memory systems [4–6] capable of rapid pattern completion through plastic recurrence. In these theories, CA3 operates as an autoassociative attractor network in which recurrent excitation stabilizes stored representations and enables pattern completion from partial cues [7–10]. While this model explains stable memory retrieval, accumulating experimental and theoretical evidence suggests that hippocampal activity also exhibits intrinsically sequential dynamics [11–16]. Rather than simply reinstating stored representations, hippocampal activity frequently unfolds as internally generated sequences, implying intrinsic temporal organization beyond static recall [17–23]. These observations raise a central question of how hippocampal circuits can maintain stable attractor-like representations while also supporting directed sequential activity.

Converging experimental evidence suggests that hippocampal population dynamics are also often inherently sequential and directional. During both sleep and awake behavior, hippocampal populations exhibit temporally ordered reactivation of place cell sequences of behaviorally relevant past and future trajectories [6,14,17,24–30]. These replay events reflect structured progression through internal representations rather than noise-driven fluctuations around a fixed point and are thought to support prediction and planning [14,15,18]. Theta phase precession, traveling waves, and directionally biased sequences further indicate that hippocampal circuits support both representational persistence and controlled temporal progression [11–13,19,22,31]. These observations motivate a closer examination of classical hippocampal recurrent network models.

In addressing this question, here the focus is on population-level modeling. While continuous-rate recurrent neural networks (RNNs) provide a tractable framework for studying low-dimensional attractor structure, they abstract away the sparse firing of hippocampal pyramidal cells. The relationship between firing-rate attractor models and low-rate spiking networks remains an open theoretical question [32–34]. However, prior work and related simulations using the population activity show that spiking networks can reproduce key qualitative dynamics of trained rate networks, including replay and phase precession, under appropriate conditions [2,35]. The present work focuses on dynamics at the population level rather than detailed spiking networks.

Classical theories of recurrent neural computation emphasize symmetry as a prerequisite for stable memory. Fully symmetric recurrent networks converge to fixed-point attractors, providing a principled mechanism for content-addressable associative memory and pattern completion [32,33,36,37]. This idea strongly influenced computational theories of hippocampal area CA3, which has long been modeled as a recurrent, symmetric autoassociative network optimized for rapid storage and retrieval of episodic patterns [4,33]. Related theoretical work has examined how specific recurrent connectivity motifs support attractor memory storage and recall, predicting structured patterns of reciprocal connectivity in networks optimized for associative memory [38]. Extensions of this approach to continuous attractor networks, including ring and bump models, demonstrated how symmetric recurrent excitation combined with broader inhibition (Mexican-hat) can stabilize low-dimensional manifolds representing continuous variables, yielding explanations for spatial coding, path integration, and head-direction representations [21,34,39–42]. Despite their success in explaining persistent activity and pattern completion, classical symmetric attractor models do not by themselves generate directed sequential dynamics. Introducing asymmetric recurrent interactions enables temporal transitions between memory states [14,39,43,44].

Early models of sequence generation in recurrent networks proposed that stable attractor dynamics can coexist with slower transition dynamics that move activity between states. For example, Sompolinsky and Kanter [44] and Kleinfeld [43] showed that combining symmetric memory couplings with asymmetric interactions or delayed synaptic response produces sequences of long-lived quasi-equilibrium states, while Lisman [14] extended this framework to hippocampal circuits by proposing a separation between autoassociative stabilization and heteroassociative progression. Related models suggested that sequences are explicitly stored through temporally asymmetric plasticity using transitions between discrete activity patterns [13,45]. Other work emphasized oscillatory mechanisms in which theta rhythms organize sequential firing and phase precession via phase coding or theta-gamma segmentation [46–49].

In addition to these mechanisms, several models have proposed that sequential dynamics arise from velocity-driven transitions along low-dimensional manifolds, in which external motion-related signals drive activity through representational space [50–52]. These approaches provide a compelling account of sequence generation during navigation and path integration. However, they rely on externally supplied velocity signals to drive transitions between states, leaving open the question of how sequential structure may be generated intrinsically within recurrent circuitry, particularly during offline states such as replay.

In contrast to these approaches which rely on transitions between discrete attractor states, externally imposed temporal structure, or transient dynamical mechanisms such as adaptation or oscillatory coupling, the present model shows that sequential dynamics can emerge intrinsically from learned recurrent connectivity. Specifically, training drives the network toward a regime in which dominant symmetric structure stabilizes a continuous attractor manifold, while a weaker, geometrically aligned antisymmetric component induces directed flow along that manifold. In this regime, replay and prediction arise as smooth, internally generated trajectories rather than discrete state transitions, without requiring explicitly stored transitions, oscillatory discretization, or hand-crafted dynamical asymmetries.

Heteroassociative chains can support sequential recall by linking successive states through asymmetric connections. However, classical analyses of CA3 networks emphasize the role of autoassociative connectivity in enabling pattern completion and stabilizing stored representations in the presence of noise or partial cues [4,14]. In these models, symmetric recurrent interactions allow the network to recover a stored representation before sequential transitions occur. Purely heteroassociative chains lack this stabilizing mechanism and can therefore be sensitive to noise and error accumulation. The regime explored here can be interpreted as a hybrid of these mechanisms, allowing stable representations and sequential dynamics to coexist within a single recurrent circuit [14,43,44]. In these models, asymmetry plays a functional role in generating sequences. However, when not structured relative to the underlying attractor geometry, increasing asymmetry can reduce the stability or persistence of individual states, leading to a tradeoff between memory stability and directed transitions. This highlights a fundamental balance: while symmetry supports memory stability, directionality is required for sequence generation, leaving open the question of how hippocampal circuits balance these competing demands. This suggests that hippocampal circuits may operate in an intermediate dynamical regime that balances stability and directed flow.

(b) The need to naturally break symmetry through neural mechanisms

Controlled symmetry breaking may link stable memory and directed temporal dynamics. In continuous attractor models, symmetric (even) recurrent interactions stabilize a low-dimensional representational manifold, while a carefully structured antisymmetric (odd) component induces smooth motion along that manifold without distorting the underlying representational structure [53]. Even modest asymmetry can convert a static attractor into a system capable of controlled flow. This approach has been used to explain head-direction integration, phase precession, and traveling population activity, in which weak antisymmetric or non-reciprocal interactions generate drift or rotation atop a stabilizing recurrent backbone [13,19,20,53–56]. Crucially, the structure of asymmetry matters: antisymmetric components aligned with spatial derivatives preserve representational structure and produce coherent translation, whereas unstructured asymmetry degrades stability or induces chaotic dynamics [44,57].

In hippocampal models, directed dynamics have been attributed to several sources of symmetry breaking, including anatomically biased recurrent connectivity, temporally asymmetric plasticity, and transient dynamical effects such as short-term synaptic processes [14,21,54,58–65]. Experimental observations are consistent this view, showing traveling hippocampal activity consistent with internally generated flow along representational axes [19,56,66]. These results highlight a central problem: hippocampal circuits must break symmetry to support replay and prediction without sacrificing stability.

One biologically plausible route to controlled symmetry breaking is synaptic plasticity. In particular, mixed symmetry in recurrent CA3 connectivity may emerge from diverse spike-timing-dependent plasticity (STDP) rules across hippocampal synapses. Early formulations of STDP emphasize a temporally asymmetric learning window in which pre-before-post activity potentiates synapses and post-before-pre activity depresses them, establishing STDP as a causality-sensitive Hebbian mechanism [67,68]. However, the CA3 may not be governed solely by canonical STDP rules.

Direct measurements at CA3-CA3 recurrent excitatory synapses reveal a largely symmetric STDP rule, well matched to CA3 autoassociative memory storage and recall [69]. Theoretical work also suggests that temporally asymmetric plasticity can coexist with more symmetric components in hippocampal-like circuits, jointly supporting pattern completion and temporal prediction [58]. Recent theory further formalizes this heterogeneity, showing that subpopulations with differing degrees of temporal asymmetry in their learning rules can regulate replay speed and directionality, with symmetric components stabilizing activity and asymmetric components driving transitions [70]. Emerging CA3 microcircuit data also indicate structural and subtype-specific asymmetries in recurrent excitation that may interact with heterogeneous plasticity to support both stable attractor dynamics and controlled sequence generation [69,71–73].

(c) Inductive biases in hippocampal RNN models

The hippocampus is frequently modeled as a recurrent neural network (RNN) due to its recurrence, plasticity, and central role in sequence learning and prediction [2,58,59,74]. A central unresolved question is whether hippocampal-like dynamics arise primarily through learning from initially unstructured connectivity, or whether circuit-level inductive biases imposed by development or evolution constrain the space of attainable dynamics. Recent RNN studies have demonstrated that predictive and sequential structure can, in principle, emerge from training networks initialized with random recurrent weights [2,54,57,74,75]. Yet because random connectivity lacks biological specificity, it remains unclear which learned features are essential for computation and which are incidental. In contrast, converging experimental and theoretical work suggests that hippocampal area CA3 is not randomly wired, but instead exhibits structured, distance-dependent recurrent connectivity that supports coordinated neural assemblies and continuous attractor dynamics [21,37,76,77]. Canonical center-surround (Mexican-hat-like) connectivity stabilizes representations through symmetric interactions, while modest symmetry breaking has long been predicted to enable directed flow [14,20,43,44,53]. These considerations motivate a systematic examination of whether biologically motivated recurrent structure shapes learning efficiency, dynamical stability, and the emergence of replay and prediction.

(d) Contribution

A central open question in hippocampal theory is how recurrent circuits stabilize internal representations while generating ordered temporal dynamics. We hypothesize that hippocampal CA3 operates in an intermediate regime where symmetric

connectivity stabilizes representations while a weaker antisymmetric component drives sequential propagation. Here, we examine how the balance between symmetric and antisymmetric recurrent connectivity governs learning, replay, and prediction in CA3-like recurrent networks.

First, using an Elman-style RNN trained on a circular prediction task inspired by place-cell navigation, we show that networks trained from random connectivity reliably self-organize into a structured regime dominated by a symmetric Mexican-hat-like component with a smaller but functionally essential antisymmetric contribution, which we term a "tilted Mexican hat." This mixed-symmetry organization emerges through learning and supports stable attractors while generating sequential dynamics underlying replay and prediction (Figure 1).

Second, we show that biologically inspired CA3-like inductive biases improve both optimization and computation (Table 1 and 2). Tilted Mexican hat initialization accelerates early learning, reinstates a two-phase optimization regime absent under random connectivity, yields tighter alignment between replay and prediction dynamics, and uncovers a spatial derivative drift component. These results show that structured recurrent connectivity does not merely emerge post hoc, but constrains learning toward functionally relevant solutions.

Third, we demonstrate that the balance between symmetric and antisymmetric connectivity constitutes a low-dimensional control parameter governing recurrent network dynamics. By explicitly manipulating the symmetry-antisymmetry ratio of structured recurrent weight matrices, we reveal transitions between static attractors under pure symmetry, unstable drift or oscillation under pure antisymmetry, and robust flow-stabilized attractors at intermediate values (Figure 2). Networks initialized near this intermediate regime learn faster, exhibit more reliable autonomous replay, and produce more accurate predictions than networks initialize with random, purely symmetric, or purely directed connectivity motifs.

Finally, we provide a biologically plausible synaptic mechanism for the emergence of mixed-symmetry connectivity. We show that a modified STDP rule combining predominantly symmetric plasticity with a weak antisymmetric component naturally produces tilted Mexican-hat connectivity profiles. This learning rule generates stable representational wave dynamics whose speed and direction depend continuously on the degree of symmetry breaking (Figure 3), linking classical STDP, recent experimental evidence for symmetric CA3-CA3 plasticity, and the emergence of directed hippocampal sequences.

Our results support a view of the hippocampus as a flow-stabilized recurrent system operating in an intermediate dynamic regime, where symmetric recurrence stabilizes internal representations and structured antisymmetry generates controlled temporal progression. *More broadly, our findings suggest that hippocampal computation reflects an interaction between task-driven learning and circuit-level inductive biases, with symmetry breaking organizing replay and prediction.*

2. Methods

(a) Task and network architecture

To examine the impact of structural inductive bias in hippocampal recurrent circuitry, we developed a predictive RNN model inspired by place-cell navigation. The task consisted of a simulated rat traversing a circular track at constant velocity. Neurons in the recurrent layer, representing CA3, received location-specific bell-shaped inputs that serve as a simplified encoding of spatial position, analogous to place fields, rather than a direct model of sensory-driven place-cell response (Fig. 1A). The network was trained to perform one-step-ahead prediction [2], mapping population activity at time t to activity at time $t+1$. The recurrent units implement population-level firing-rate dynamics and do not explicitly model spike timing, synaptic conductances, or oscillatory interactions present in biological CA3 circuits. Throughout, the recurrent weight matrix is interpreted as an effective coupling in a reduced rate model. Because Dale's principle is not enforced, negative weights represent effective net inhibitory influence at the population level.

The model was implemented as an Elman RNN in PyTorch (v2.6.0) and trained using gradient descent with backpropagation through time and a fixed learning rate. Recurrent units used a $\tanh$ nonlinearity, chosen for its bounded output and smooth derivative which facilitate stable training and recurrent dynamical analysis, and output units used a sigmoid nonlinearity. Input-to-hidden and hidden-to-output weights were frozen to isolate the role of recurrent dynamics. Training terminated when either a maximum epoch count was reached or the loss dropped below 1% of its initial value and changed by less than $10^{-3}\%$ over ten consecutive epochs. The network was first trained starting from a random initialization then using crafted structural hidden weights. The use of inductive bias for improved training was motivated by pretraining (see Fig. S1), but the hidden weights were initialized from different connectivity families. Network dimensions and hyperparameters are summarized in Table S1. Training sequences consisted of repeated laps around the circular track with fixed direction and velocity.

(b) Recurrent connectivity families and inductive biases

We examined multiple recurrent connectivity families encoding distinct inductive biases: random, identity, cyclic shift, and Mexican-hat connectivity. Random initialization served as the baseline with other families normalized to match.

We explicitly define the connectivity families as follows. Let $W \in \mathbb{R}$ denote the recurrent weight matrix.

Identity connectivity is defined as

$$W_{ij} = \delta_{ij} \tag{2.1}$$

where δ_{ij} is the Kronecker delta. This corresponds to purely self-recurrent dynamics with no interactions, thereby isolating temporal persistence.

Cyclic shift connectivity is defined as

$$W_{ij} = \delta_{i,(j+1) \pmod N} \tag{2.2}$$

which implements a discrete translation operator on the ring, such that activity at neuron j is passed to neuron $j+1$ with wrap-around. Although visually similar to identity (a diagonal matrix), cyclic shift connectivity is strictly off-diagonal, implementing directional propagation rather than persistence.

Both the cyclic shift and Mexican-hat matrices are circulant, meaning their rows are cyclic permutations of a base kernel, ensuring translation-invariant connectivity on the ring.

Further discussion and visualizations of all connectivity families (random, identity, cyclic shift, and Mexican-hat) are provided in the Supplemental Material 6(c) (see Fig. S2).

(c) Symmetry-antisymmetry decomposition and control parameter

In addition to the baseline cyclic shift and Mexican-hat initializations, we also implemented variants with predetermined symmetry components. Specifically, the recurrent matrices W were decomposed into symmetric S and antisymmetric A components

$$S = \frac{1}{2}\left(W + W^T\right) \quad A = \frac{1}{2}\left(W - W^T\right) \tag{2.3}$$

and recombined as

$$W(\alpha_0) = (\alpha_0)S + (1-\alpha_0)A \quad \text{with } \alpha_0 \in [0,1] \tag{2.4}$$

The parameter α_0 controls the initial balance between reciprocal (symmetric, $\alpha_0 = 1$) and directional (antisymmetric, $\alpha_0 = 0$) interactions. To track how symmetry evolved during training, we defined a time-dependent symmetry index using Frobenius norms as

$$\alpha(t) = \frac{||S||_F}{||S||_F + ||A||_F} \quad \alpha \in [0,1] \tag{2.5}$$

which provides a low-dimensional summary of recurrent structure, explicitly parameterizing the degree of symmetry breaking in the network. α increases as the symmetric component becomes dominant, where $\alpha = 0$ represents the purely anti-symmetric extreme and $\alpha = 1$ denotes the purely symmetric state.

(d) Training, replay, and prediction protocols

Networks were trained using teacher forcing [78], with the true input provided at each time step. We used two evaluation regimes: prediction and replay. For replay, networks were driven entirely by low-amplitude noise to assess internally generated dynamics. For prediction, networks received a short prefix of true input followed by noise and were assessed on their ability to autonomously continue the learned sequence.

(e) Mexican-hat connectivity construction

Mexican-hat connectivity was constructed by defining a one-dimensional Difference-of-Gaussians (DoG) kernel on a ring of size N. For offsets from the diagonal $d \in 0, 1, ..., H-1$, we define the signed circular distance as

$$d_{signed} = ((d + \lfloor H/2 \rfloor) \mod H) - \lfloor H/2 \rfloor, \quad r(d) = |d_{signed}| \tag{2.6}$$

and construct the DoG kernel

$$\mathrm{DoG}(d) = a \exp\left(-\frac{r(d)^2}{2\sigma_e^2}\right) - b \exp\left(-\frac{r(d)^2}{2\sigma_i^2}\right) \tag{2.7}$$

where σ_e and σ_i are the excitation and inhibition widths, and a and b are the excitatory and inhibitory amplitude respectively. b was chosen to approximately balance total excitation and inhibition. The DoG inhibitory lobe denotes an effective negative interaction in the rate model. Enforcing Dale's principle would require an explicit excitatory-inhibitory architecture or sign-constrained parameterization, which is not imposed here. In rate-based network models, the recurrent weight matrix should be interpreted as an effective interaction operator rather than a direct representation of individual synaptic connections. The kernel was scaled to match a target spectral radius ρ_{target} using the maximum magnitude of its Fourier coefficients. In our experiments, $\sigma_i > \sigma_e$ to produce a local excitatory bump with broader surround inhibition.

The final recurrent weight matrix was obtained by embedding the kernel into a circulant matrix

$$W_{ij} = \mathrm{DoG}([(i - j - \Delta) \bmod N]) \tag{2.8}$$

where Δ optionally introduces a directional shift. For shifted Mexican-hat matrices, the symmetry-antisymmetry balance was further controlled via the α decomposition described above.

(f) Plasticity analysis

To identify a local synaptic mechanism capable of producing the mixed-symmetry connectivity for hippocampal sequence generation, we developed a modified spike-timing-dependent plasticity (STDP) rule [67]. This STDP rule completely replaces backpropagation in the final experiment. While canonical STDP is temporally asymmetric, strengthening synapses only when presynaptic activity precedes postsynaptic firing [68],this causal rule contradicts the dual role of the hippocampus in forming stable memory representations and enabling directed sequential dynamics.

Our rule is motivated by recent empirical evidence suggesting that CA3 recurrent circuitry is not governed solely by canonical, asymmetric STDP [69]. Direct electrophysiological measurements at CA3-CA3 recurrent excitatory synapses have revealed a predominantly symmetric STDP rule, where potentiation occurs for both pre-before-post and post-before-pre spike timings. This symmetric plasticity is thought to be functionally optimized for the rapid storage and stable retrieval of autoassociative memories. We therefore implement a rule that allows for the strengthening of "reverse" synapses (post-before-pre) during forward spiking events, effectively modeling the retroactive potentiation observed in biological CA3 ensembles.

We formalize this by modifying the canonical STDP such that it linearly interpolates between a standard antisymmetric STDP kernel, $K(s)$, and a retroactive, symmetry-inducing component controlled by a scalar parameter $\epsilon \in [0, 1]$. The total synaptic update from neuron j to neuron i is calculated by integrating the product of neural activities $u(t)$ across a biphasic kernel:

$$W_{ij} = \int_{-\infty}^{\infty} K(s) \int_{-\infty}^{\infty} u_i(t) u_j(t-s) dt ds + \epsilon \int_{-\infty}^{\infty} K(s) \int_{-\infty}^{\infty} u_j(t) u_i(t-s) dt ds \tag{2.9}$$

where the biphasic STDP kernel $K(s)$ is defined as:

$$K(s) = \begin{cases} A_+ e^{-s/\tau_+} & \text{if } s \geq 0 \text{ (LTP)} \\ -A_- e^{s/\tau_-} & \text{if } s < 0 \text{ (LTD)} \end{cases} \tag{2.10}$$

In these expressions, A_+ and A_- represent the maximum learning rates, while τ_+ and τ_- denote the time constants for potentiation and depression, respectively. The parameter ϵ serves as a control for the degree of symmetry breaking at the synaptic level. This formulation allows the circuit to interpolate between purely antisymmetric ($\epsilon = 0$) and purely symmetric ($\epsilon = 1$) regimes, providing an analytically tractable mechanism for the hippocampus to balance representational stability with controlled temporal progression through local changes in the retrograde plasticity signal. We then analyzed the emergent network structure when a simple linear RNN is stimulated with the moving Gaussian input activity profile of the form used in our experiments.

3. Results

(a) Task-driven symmetry breaking in recurrent connectivity supports replay and prediction

Training an Elman recurrent neural network from random initialization was sufficient to generate hippocampal-like replay and one-step-ahead prediction on a circular navigation task (Fig. 1A-C). This is consistent with prior work showing that generic recurrent networks can support sequential dynamics under appropriate task constraints [2,54]. We therefore focus on the structure of the learned recurrent connectivity that gives rise to these dynamics.

Despite starting from unstructured random weights, learning consistently reshaped the recurrent matrix into a highly organized form, indicating that the task selects this structure rather than imposing it architecturally. When neurons were sorted by their peak activation time during the training sequence, the final recurrent connectivity exhibited a clear banded organization (Fig. 1D). Averaging weights along the diagonals revealed a Mexican-hat-like interaction profile characterized by strong local excitation and longer-range inhibition. Negative couplings reflect effective inhibition in the reduced model. Such profiles are canonical in attractor models and support low-dimensional ring manifolds capable of stable pattern completion [21,32–34,36,37,39,40,79].

Crucially, the learned connectivity was not purely symmetric. The eigenspectrum of the recurrent matrix, W_{hh}, contained complex-valued modes (Fig. 1D, right), indicating the presence of antisymmetric structure. While most eigenvalues clustered near the origin, a small number of dominant modes emerged. In particular, the leading eigenvalue had a real component of 2.16 and a nonzero imaginary component of $\pm$ 0.15, indicating a dominant recurrent mode with weak rotational structure. Because these spectra are computed from the recurrent weight matrix W_{hh}, this does not imply instability of the nonlinear dynamics. In a $\tanh$ RNN, the effective Jacobian is $J_t = \mathrm{diag}(1 - h_t^2) W_{hh}$, so the nonlinear gain contracts the spectrum during both replay and prediction. Further analysis of the effective gain can be found in Supplemental Material 6(d).

To isolate the contributions of these components, we decomposed the learned recurrent matrix into symmetric and antisymmetric parts (Fig. 1E-F). The symmetric component dominated the overall structure, as measured by its Frobenius norm

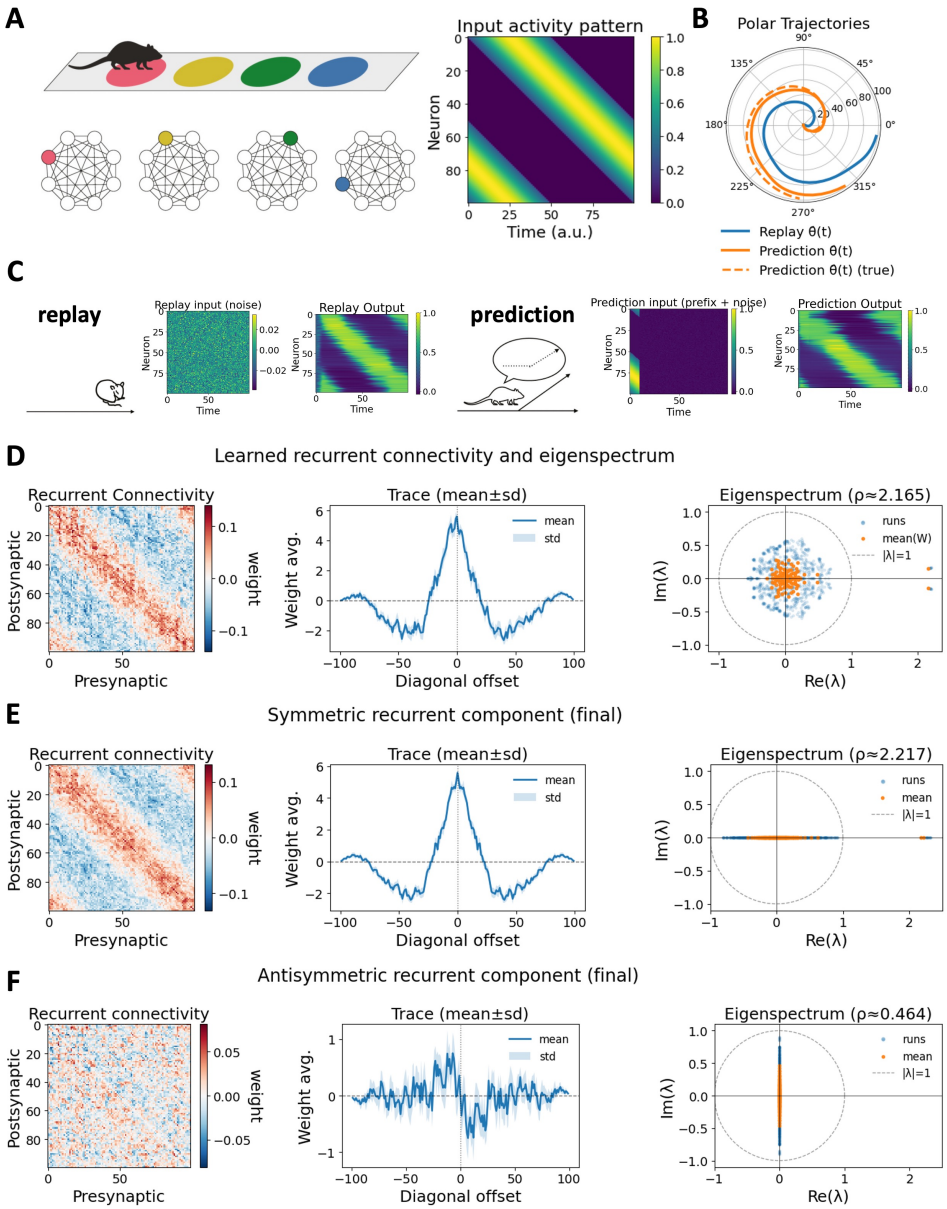

Figure 1: **Learned symmetry breaking in recurrent connectivity supports replay and prediction. (A)** Task schematic: a rat traverses a circular track and receives bell-shaped, location-specific input. **(B)** Polar trajectories during replay (*blue*) and prediction (*orange*), showing coherent sequential dynamics. **(C)** Evaluation paradigms: replay driven by noise (*left*) and prediction following a brief input prefix (*right*). **(D)** Learned recurrent weight matrix, averaged across runs and sorted by neurons' peak activation time, revealing a banded structure aligned with sequence order. **(E)** Symmetric component of the learned connectivity, exhibiting a Mexican-hat-like profile and real-valued eigenmodes. **(F)** Antisymmetric component, showing off-diagonal structure and imaginary eigenmodes consistent with directional or flow-like dynamics. Eigenspectra in panels D-F are computed from the recurrent weight matrix W_{hh} and are interpreted as structural diagnostics of learned connectivity rather than direct stability criteria for the nonlinear dynamics.

($||W_{sym}||_F = 3.65$), which quantifies the total magnitude of the recurrent weights. It reproduced both the fully symmetric Mexican-hat trace profile and a purely real eigenspectrum (Fig. 1E), consistent with stable fixed-point and continuous attractor dynamics [21,36,37,79]. Although these dynamics resemble those of continuous attractor models, we do not claim that the trained network implements a mathematically exact continuous attractor. In contrast, the antisymmetric component, though smaller in weight magnitude ($||W_{anti}||_F = 1.94$), accounted for approximately 35% of the total recurrent norm and exhibited imaginary eigenmodes within the unit circle (Fig. 1F), consistent with rotational or flow-like dynamics that bias activity propagation along the ring [53].

The spatial structure of the antisymmetric component closely resembles a noisy first spatial derivative of the symmetric Mexican-hat profile, with odd symmetry around zero offset and localization near the attractor peak (Fig. 1F, middle). Such derivative-like antisymmetric perturbations have been shown to convert static continuous attractors into dynamic solutions by introducing a controlled direction bias [53].

This decomposition shows how replay and prediction emerge from the learned connectivity. The symmetric component stabilizes representations on a low-dimensional attractor-like manifold, while the antisymmetric component breaks time-reversal symmetry and selects flow direction along that manifold. The resulting dynamics correspond to activity moving along a ring manifold, producing smooth phase advance in polar coordinates and coherent sequential structure during both replay and prediction (Fig. 1B). These trajectories do not perfectly overlap, indicating that the learned flow is flexible and not exactly aligned with the task manifold. The dominant outlier eigenmode combines strong symmetric gain with a weak antisymmetric contribution, suggesting that sequence generation arises from a mode that is gently rotated in state space rather than purely oscillatory dynamics [53,54,58,59].

These results demonstrate that optimization for one-step-ahead prediction from random connectivity reliably induces a mixed-symmetry recurrent organization, rather than converging to a purely symmetric attractor network. This provides evidence that symmetry breaking can be learned from task demands, while the residual mismatch between replay and prediction dynamics suggests that learning alone does not fully constrain the recurrent structure. These findings show that mixed-symmetry can emerge from task optimization alone, motivating a systematic examination of how structured recurrent inductive biases shape learning and performance.

(b) CA3-like recurrent inductive bias improves replay, prediction, and learning dynamics

To assess how recurrent inductive bias shapes hippocampal-like computation, we compared networks initialized with different CA3-inspired recurrent motifs (identity, cyclic shift, Mexican-hat, see Methods section 2(b)) and evaluated autonomous replay and one-step-ahead prediction. Replay probes internally generated dynamics without informative input, whereas prediction tests whether learned temporal structure is maintained after withdrawal of external drive. Performance across both replay and prediction depended strongly on the initial recurrent structure. Networks initialized with a constructed Difference-of-Gaussians Mexican-hat (center-surround) connectivity (see Methods section 2(e)) consistently achieved the strongest performance across all metrics (Tables 1-2; Fig. 2).

(i) Replay

During replay, Mexican-hat initialization with an initial symmetry value of $\alpha_0 = 0.70$ yielded the highest ring-decode R^2 and near-maximal residual lag-1 autocorrelation (Table 1), indicating that internally generated activity remains confined to the learned ring manifold while evolving smoothly over time. More information on metrics used can be found in the Supplemental Material section 6(e).

Table 1: Replay metrics

metric	random	identity	pure shift	mixed shift	Mexican hat ($\alpha_0 = 0.7$)
ring decode R^2 (↑)	0.963 ± 0.012	0.211 ± 0.27	0.782 ± 0.013	0.929 ± 0.027	**0.975** ± 0.0029
residual lag-1 autocorr (↑)	0.99 ± 0.0032	0.176 ± 0.27	0.99 ± 0.0029	0.99 ± 0.002	**0.99** ± 0.0061

Random initialization also supported replay, showing that task-driven learning alone can induce continuous-attractor-like dynamics. However, replay trajectories were less consistent across runs and showed weaker alignment between replay and prediction than in Mexican-hat networks (Fig. 2D). In contrast, identity initialization, which is maximally symmetric and lacking lateral interactions, failed to support replay as shown by exhibiting low ring-decode R^2 and poor temporal coherence. Thus, temporal persistence alone is insufficient for autonomous sequence generation. Further results from the identity and cyclic shift matrices can be found in the Supplemental Material section 6(c).

(ii) Prediction

Prediction performance further separated the connectivity families (Table 2). Mexican-hat networks simultaneously maximized angular accuracy (ring-decode R^2 and mean resultant length R) while minimizing mean squared error. Cyclic shift connectivity, which hardwires directionality, improved prediction and replay relative to identity initialization but underperformed Mexican-hat connectivity and exhibited replay breakdown at higher symmetry levels (Supplementary Figs. S4-S5). Thus, directionality alone is also insufficient without a stabilizing recurrent scaffold.

Table 2: Prediction metrics

metric	random	identity	pure shift	mixed shift	Mexican hat ($\alpha_0 = 0.7$)
mean resultant length R (↑)	0.956 ± 0.0071	0.939 ± 0.015	0.836 ± 0.0083	0.922 ± 0.024	**0.987**± 0.0044
ring decode R^2 (↑)	0.968 ± 0.011	0.957 ± 0.014	0.778 ± 0.012	0.921 ± 0.036	**0.99**± 0.0039
MSE (↓)	0.0646 ± 0.0066	0.079 ± 0.0076	0.114 ± 0.0052	0.0814 ± 0.013	**0.0499**± 0.0046

Structured inductive bias also altered optimization dynamics. The constructed Mexican-hat initialization often began at lower loss and exhibited a pronounced early reduction in error compared to random connectivity (Fig. 2A,B). For specific initial symmetry values, training showed a two-phase learning profile that was absent under random initialization, with an early rapid

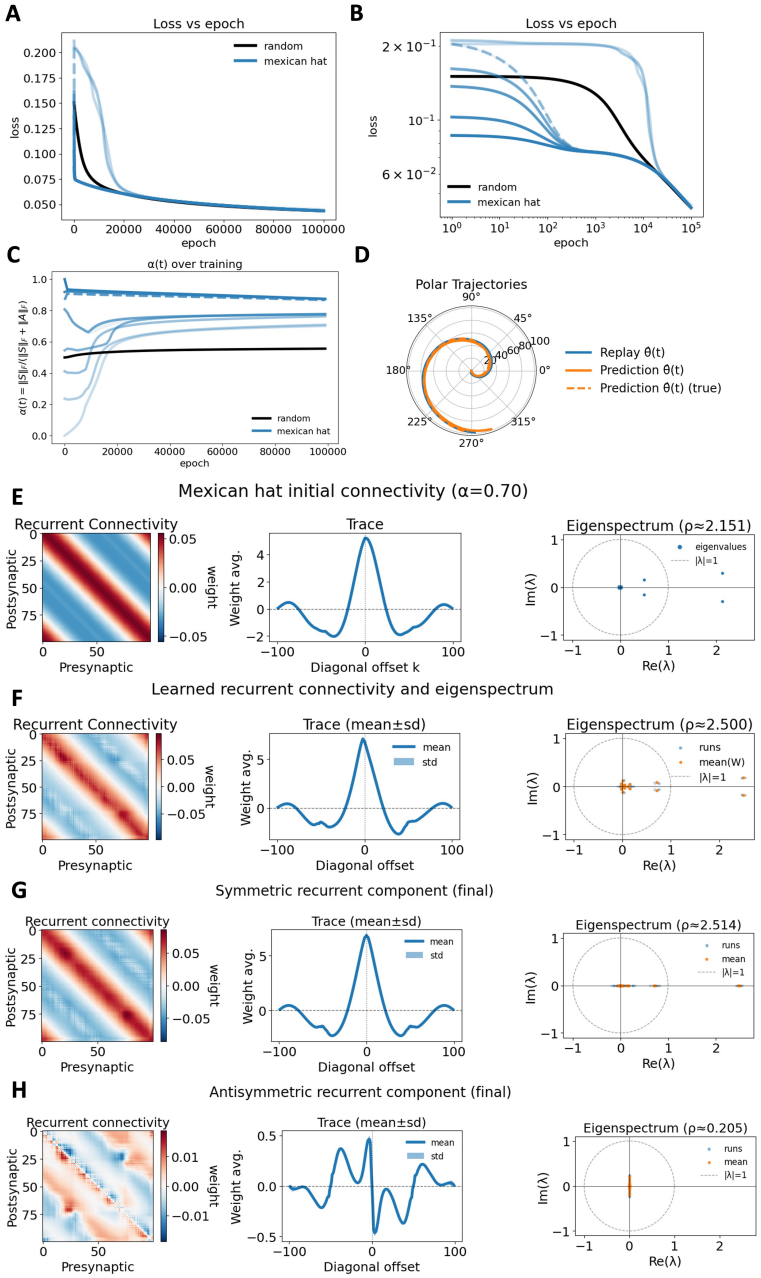

Figure 2: **Tilted Mexican-hat CA3-like recurrent connectivity accelerates learning and stabilizes sequential dynamics. (A,B)** Training loss for networks initialized with circulant Mexican-hat connectivity (blue; varying initial symmetry α_0, dashed: $\alpha_0 = 0.70$) compared with random connectivity (black), shown on linear and log-log axes. **(C)** Evolution of the symmetry index $\alpha(t)$, showing convergence to an intermediate, symmetry-dominated regime. Initial symmetry α_0 is given by value at epoch 0. **(D)** Polar trajectories of replay and one-step-ahead prediction for $\alpha_0 = 0.70$. **(E)** Initial Mexican-hat connectivity ($\alpha_0 = 0.70$), its diagonal trace, and eigenspectrum. **(F)** Learned recurrent connectivity after training. **(G,H)** Symmetric (G) and antisymmetric (H) components of the learned connectivity, showing a dominant attractor-like scaffold with a weaker directional component. Eigenspectra shown in panels E-H correspond to W_{hh}. The effective Jacobian is contracted by the state-dependent $\tanh$ gain during replay and prediction.

loss decrease followed by slower refinement [80,81]. The two-phase learning dynamics suggest that networks first learn the input-output mapping and subsequently refine the recurrent structure that stabilizes and directs activity along the task manifold. This indicates that structured recurrence not only improves final performance but also places the network in a more favorable regime for learning. These advantages raise the question of what recurrent structure is ultimately selected by optimization under different initial conditions. We therefore next examine how learning reshapes recurrent connectivity and whether it converges to a common dynamical regime across structured initializations.

(c) Learning converges to a mixed-symmetry regime supporting sequential dynamics

Across structured initializations, learning consistently drove recurrent connectivity toward an intermediate balance between symmetric and antisymmetric components, quantified by the symmetry index $\alpha(t)$ (Fig. 2C). Despite wide variation in initial symmetry, $\alpha(t)$ converged to a predominantly symmetric regime with a persistent antisymmetric component, indicating that intermediate symmetry is task-selected rather than incidental. Across all conditions, the best performance was obtained for off-centered Mexican-hat connectivity initialized with an intermediate symmetry level ($\alpha_0 = 0.70$), which places the network near the task-selected symmetry regime from the outset (Fig. 2A-C).

The learned recurrent weights preserve a Mexican-hat-like center-surround profile characterized by strong local excitation and broader inhibition (Fig. 2E-F). Decomposition of the final connectivity revealed that the symmetric component dominated the overall structure and eigenspectrum (Fig. 2G, $||W_{sym}||_F = 3.7$), consistent with stabilization of an attractor manifold. In contrast, the antisymmetric component was smaller in magnitude (Fig. 2H; $||W_{anti}||_F = 0.42$, $||A||/(||S||+||A||) = 0.10$) but introduced imaginary eigenmodes that bias activity propagation. As in the randomly initialized networks, the learned antisymmetric component closely resembles a noisy spatial derivative of the symmetric Mexican-hat profile, with odd symmetry around zero and localization near the attractor peak (Fig. 2H, middle). This derivative-like structure is consistent with theoretical predictions that weak antisymmetric perturbations of attractor-like manifolds induce directional drift without destabilizing the underlying manifold [53].

Functionally, this mixed-symmetry regime supports dynamics that are neither purely static nor purely feedforward. Replay and prediction trajectories were more closely aligned for Mexican-hat initializations than for random connectivity (Fig. 2D), indicating that internally generated dynamics reflect the learned predictive structure. Spectral analyses further showed that the best-performing initializations already contained weak imaginary components (Fig. 2E), which were preserved and refined during training (Fig. 2F-H).

These results show that task optimization reliably selects an intermediate symmetry regime in which symmetric recurrence stabilizes representations while structured antisymmetry enables controlled temporal progression. Task learning can recover mixed-symmetry recurrent structure from random connectivity, but structured inductive bias accelerates learning, stabilizes dynamics, and yields the most robust replay and prediction. In the next section, we examine how a biologically plausible plasticity mechanism, specifically a modified STDP rule, can give rise to this structured symmetry breaking.

(d) Heterogeneous STDP induces symmetry breaking in Mexican-hat connectivity

To connect the mixed-symmetry structure observed in the trained RNNs to biologically plausible learning mechanisms, we analyzed the RNN structure emergent from a modified STDP rule that interpolates between antisymmetric and symmetric plasticity. This modified STDP rule replaces backpropagation used in the previous experimental setups. Classical STDP is temporally asymmetric, strengthening synapses when presynaptic activity precedes postsynaptic firing and weakening the reverse. While such causal learning supports sequence formation, it does not by itself produce the stabilized center-surround structure observed in the learned recurrent connectivity (Figure 2). Conversely, purely symmetric Hebbian plasticity produces stable attractors but lacks directional bias [36].

Under a traveling Gaussian input, the modified STDP learning rule produces a family of synaptic weights whose spatial interaction pattern is determined by the symmetry parameter ϵ (see Methods section 2(f) and analytical derivation in Supplementary section 6(f); Fig. 3). For $\epsilon = 0$, the rule reduces to standard STDP and yields a strongly antisymmetric interaction profile, dominated by imaginary eigenvalues and fast drift of the activity bump (Fig. 3A). For $\epsilon = 1$, a symmetric Mexican-hat profile is produced with a purely real eigenspectrum (Fig. 3B). For intermediate values ($\epsilon = 0.85$), the learned synaptic weight profile becomes a weakly tilted Mexican hat, preserving the stabilizing center-surround structure while introducing a small antisymmetric component (Fig. 3C). The corresponding eigenspectrum contains dominant real modes with only small imaginary components, yielding a stable activity bump that propagates at a constant velocity.

Although ϵ plays a role analogous to the symmetry index α, the two are not identical. ϵ parametrizes symmetry at the level of synaptic plasticity, whereas α quantifies the symmetry of the overall connectivity. Nevertheless, intermediate values of ϵ naturally give rise to recurrent connectivity in the same mixed-symmetry regime favored by task optimization in the RNNs.

The analysis of the modified STDP learning rule shows that the flow-stabilized attractor regime is not merely an outcome of backpropagation but is also accessible to biological circuits via simple tuning of the retrograde plasticity signal (ϵ). In biological terms, ϵ may correspond to the strength of retroactive potentiation observed in recent STDP experiments in CA3 [69] allowing the hippocampus to interpolate between storage (attractor) and replay (flow) modes through purely local changes.

4. Discussion

(a) Symmetry breaking as a mechanism for hippocampal sequence generation

We have built on previous work to provide a quantitative analysis of partial symmetry breaking in recurrent connectivity as a circuit-level principle that enables hippocampal networks to jointly support stable memory representations, internally generated replay, and predictive sequence dynamics. Using recurrent neural networks trained on a prediction task, we show that these functions emerge most robustly in an intermediate regime dominated by symmetric structure but augmented by a weaker

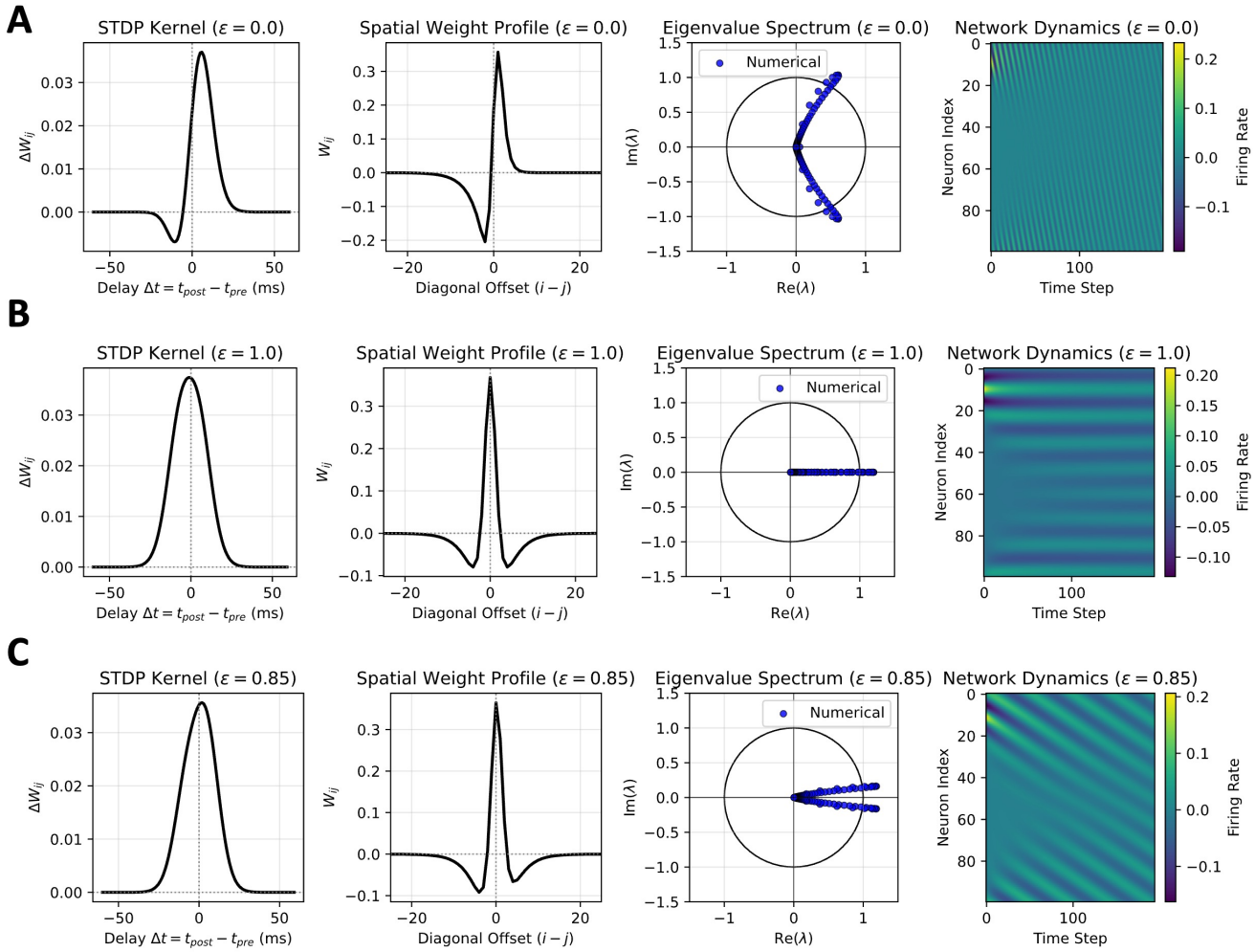

Figure 3: **Symmetry breaking in STDP generates flow-stabilized Mexican-hat dynamics.** Each panel shows the learned synaptic weight profile for different settings of the symmetry parameter (ϵ). The first column shows the STDP learning kernel, the second column the weight profile after learning, the third column the computed eigenvalue spectrum of the learned matrix and the fourth column the resulting network activity dynamics. We use STDP with parameters $A_+ = 1.0, A_- = 0.5, \tau_+ = 1.0, \tau_- = 3.0$. **(A) Pure symmetry** ($\epsilon = 1$): symmetric Mexican-hat profile with real eigenvalues and a stationary attractor. Activity bump is stationary (zero velocity). **(B) Pure antisymmetry** ($\epsilon = 0$): standard STDP gives highly asymmetric weight profile with fast activity drift. **(C) Symmetry breaking** ($\epsilon = 0.85$): weakly tilted Mexican hat with mixed real-imaginary eigenvalues, producing a flow-stabilized attractor with a slowly propagating activity bump.

antisymmetric component. This mixed-symmetry organization stabilizes internal representations while enabling directed flow along them, yielding dynamics that unify replay and prediction within a single recurrent circuit.

This regime is not imposed but selected by learning. Across random and structured initializations, networks reliably converge toward a tilted Mexican-hat-like connectivity, suggesting that symmetry breaking acts as a low-dimensional control parameter governing the balance between stability and motion. These findings support a view in which hippocampal circuitry is shaped by biological inductive biases that constrain learning toward a family of dynamical regimes well suited for sequential memory, with experience refining task-specific operating points within that space rather than discovering structure *de novo*.

Symmetric recurrent connectivity stabilizes attractor states but suppresses temporal structure, whereas strong asymmetry introduces directionality at the expense of stability. A long line of theoretical work has shown that symmetry breaking in recurrent networks is essential for generating ordered temporal dynamics [13,20,41,53,61]. In hippocampal models, asymmetric interactions have been invoked to explain path integration, sequential recall, and internally generated trajectories [14,21,54,58,60].

Our results extend this framework by showing that replay and prediction emerge most robustly when symmetric and antisymmetric components coexist in a controlled ratio. Networks trained from various initial conditions reliably converge toward this mixed-symmetry regime, suggesting that partial symmetry breaking is a task-selected dynamical solution.

(b) Circulant connectivity and flow-equivariant dynamics in hippocampal networks

An important theoretical insight emerging from this work is the role of translation-invariant recurrent structure, formalized as Toeplitz or, more specifically, circulant connectivity [55,82]. In such networks, synaptic weights depend primarily on relative position, yielding spatial Fourier eigenmodes. Symmetric Toeplitz structure selectively stabilizes low-frequency modes, producing smooth continuous attractors, whereas antisymmetric components introduce imaginary eigenvalues that generate coherent drift along the attractor manifold.

From this perspective, the balance between symmetric and antisymmetric structure directly determines whether dominant modes are stationary or propagating. Our findings show that learning consistently drives recurrent connectivity toward this structured regime, even when starting from random matrices. This suggests that circulant-like structure is not merely analytically convenient but an efficient solution for predictive sequence learning.

Approximate translation invariance is biologically plausible in hippocampal CA3, where recurrent connectivity exists and lacks obvious spatial anchoring. Experience-dependent plasticity would naturally reinforce relative rather than absolute coding, while deviations from perfect symmetry provide a principled mechanism for generating directed flow without disrupting representational stability. In this way, partial symmetry breaking transforms a static attractor into a flow-stabilized representation.

Our model provides a concrete instantiation of this principle: partial symmetry breaking within a circulant recurrent scaffold yields dynamics that unify replay and prediction through flow along an internal representational manifold. In this sense, hippocampal recurrent dynamics implement a form of flow-equivariant computation [82], in which task-relevant structure is preserved under continuous transformations while temporal progression is encoded through controlled motion along that structure.

(c) Relationship to prior models

Zhang's seminal work explicitly decomposed recurrent connectivity into even (symmetric) and odd (antisymmetric) components, and demonstrated that the odd component generates rotational flow in head-direction networks [53]. He further showed that optimal dynamics arise in an intermediate regime dominated by symmetry but containing a weaker antisymmetric component, and that the antisymmetric structure approximates the spatial derivative of the symmetric component.

Our results closely parallel this model but extend it in several key ways. First, whereas Zhang prescribed the relative strength of even and odd components to implement velocity integration, we show that this balance can emerge through learning under a predictive objective. Second, Zhang's model relies on externally driven velocity inputs, whereas our networks generate replay and prediction internally, in the absence of explicit motion signals. Finally, although the learned antisymmetric component closely resembles the spatial derivative of the symmetric Mexican-hat profile, this relationship arises naturally from optimization rather than being imposed. Together, these findings extend symmetry-based theories from hand-designed integrators to learned hippocampal dynamics.

These results can be viewed as a continuous generalization of earlier fast-slow attractor models [13,14,43–45], in which stabilization and transition dynamics were implemented as separate processes, showing instead that both can emerge within a single recurrent circuit through learned partial symmetry breaking.

Recent theoretical syntheses emphasize that neural circuits often operate between pure attractor and pure integrator regimes [83]. Our results identify partial symmetry breaking as a concrete mechanism for realizing such intermediate dynamics in hippocampal networks.

(d) Structured versus random recurrent connectivity

Random recurrent networks can exhibit rich dynamics, particularly near the edge of chaos [57], and training can harness this richness for computation [75,84]. However, such networks often require fine-tuning to suppress instability and may lack robustness across tasks.

In contrast, we find that structured inductive biases, specifically Mexican-hat-like connectivity with partial symmetry breaking, provide a stable substrate for learning. Networks initialized near this regime learn faster, converge more reliably, and exhibit more interpretable dynamics than those initialized randomly. This aligns with work showing that population structure constrains and stabilizes computation in recurrent circuits [74,85]. From a biological standpoint, these results support the idea that hippocampal circuitry may be developmentally biased toward certain connectivity motifs, with learning refining rather than discovering them *de novo*.

(e) Biological plausibility

The coexistence of symmetric and asymmetric connectivity need not imply multiple specialized learning rules. Synapses exhibit substantial heterogeneity in timing, plasticity, and short-term dynamics [58,69,71,86]. Our heterogeneous STDP results demonstrate that realistic synaptic diversity can naturally give rise to the structured symmetry breaking required for flow, providing a plausible biological substrate for the dynamics observed here.

The relationship between firing-rate attractor models and biologically realistic spiking networks is an important consideration. CA3 pyramidal neurons fire sparsely, raising the question of how smooth attractor dynamics emerge from low-rate, temporally

discrete spike trains. The present model does not attempt to directly resolve this gap, instead adopting a mesoscopic, population-rate description in which recurrent symmetry breaking shapes effective dynamical flow. However, prior work has shown that trained rate networks can be mapped to spiking implementations that preserve key dynamical features [35]. Consistent with this, applying this mapping to the present model yields replay and theta phase precession in a network of leaky integrate-and-fire neurons [2], suggesting that the learned dynamical structure can, in some cases, be expressed in a spiking regime. That said, whether this specific flow-stabilized attractor can be robustly implemented in sparse spiking networks remains an important direction for future work.

The Mexican-hat profiles reported here represent effective interaction at the level of population dynamics. In biological CA3 circuits, a comparable center-surround structure would arise through excitatory-inhibitory microcircuit organization. Several studies have demonstrated that recurrent networks respecting Dale's principle can reproduce similar dynamical regimes by explicitly separating excitatory and inhibitory populations [87–91]. The present study focuses on the dynamical consequences of symmetry structure in the effective recurrent operator, leaving the detailed microcircuit implementation to future work.

Experimental measurements suggest that CA3 recurrent excitatory connectivity is sparse and only weakly distance-dependence across several species [72,77,92–94]. The Mexican-hat profile used here should therefore be interpreted as an effective population-level connectivity motif rather than a direct anatomical description. In hippocampal circuits, structured inhibition, heterogeneous plasticity rules, and network dynamics can shape the interactions experienced by pyramidal populations, potentially giving rise to center-surround structure at the population level despite sparse underlying connectivity. An important direction for future work is to determine whether the mixed-symmetry regime identified here persists in biologically constrained architectures incorporating sparse connectivity and Dale's principle.

An important distinction is that the traveling waves observed in our model are representational rather than explicitly anatomical. However, these two forms of dynamics may be related. Anatomical traveling waves have been observed along the septotemporal (dorsal-ventral) axis of the hippocampus, demonstrating that activity can propagate across this axis in a coordinated manner [19,66]. At the same time, place fields are spatially distributed and partially redundant, such that neurons representing nearby or even similar locations are not arranged topographically but instead interspersed across the network [8,95–98]. In such a distributed code, a propagating anatomical wave need not correspond to a contiguous trajectory in physical space. Instead, as activity moves along the dorsal-ventral axis, it could sequentially recruit neurons whose place fields collectively form a trajectory through the environment, even though those neurons are spatially interspersed within the hippocampus. Thus, the structured propagation required for anatomical traveling waves could arise from an underlying recurrent connectivity motif with local excitation and longer-range inhibition, qualitatively consistent with a tilted Mexican-hat profile. We therefore speculate that the same class of connectivity that supports anatomical wave propagation may also give rise to the internally generated representational sequences observed during replay and prediction. Importantly, this interpretation does not imply that anatomical waves directly encode spatial trajectories, but rather that they reflect a more general dynamical principle that structured recurrent interactions can generate propagating activity patterns which, when combined with a distributed spatial code, appear as flexible sequences in representational space. This hypothesis makes a testable prediction that, if correct, connectomic analyses of CA3 should reveal effective interactions consistent with a weakly asymmetric (tilted) center-surround structure along the septotemporal axis.

Our findings support a view of hippocampal CA3 as a flow-stabilized recurrent circuit in which memory, replay, and prediction emerge from a balance between symmetric stabilization and directed flow. Conceptual frameworks emphasize that the hippocampus is fundamentally a sequence generator, with internally generated, ordered dynamics forming the basis of memory and prediction rather than static representations [3,12,13,19,22,31,99]. We find that recurrent networks trained on predictive sequence tasks converge to a regime dominated by Mexican-hat-like symmetric connectivity with a smaller but essential antisymmetric component. This mixed-symmetry organization is consistent with experimental observations in CA3, where recurrent synapses exhibit largely symmetric plasticity optimized for autoassociative storage [69], alongside structural and functional asymmetries between pyramidal subtypes that support sequential activation [71–73,77]. The CA3 is difficult to directly measure connectivity, but more generally, connectomic studies show strongly directed and largely acyclic connectivity in cortical microcircuits, indicating biological networks can display structural asymmetry [100]. Although learning from random connectivity can recover this regime, structured CA3-like inductive biases accelerate convergence and stabilize dynamics. This suggests that hippocampal circuitry constrains learning toward dynamical regimes well suited for sequential memory tasks.

This framework yields clear experimental predictions. Selectively perturbing directional coupling should impair replay speed or directionality without abolishing completion (e.g. targeted optogenetic manipulation of CA3 ensembles [101], while state-dependent modulation of CA3 input pathways should bias internal flow without disrupting the underlying attractor geometry [86]. While the exact symmetry level, propagation speed, and gain depend on task demands, the principle of dominant symmetry augmented by weak derivative-like antisymmetry appears to be task-general. These findings argue that hippocampal sequence generation reflects an interaction between biological inductive biases and task-driven learning, rather than learning alone.

These results suggest that hippocampal replay and prediction emerge from a delicate balance between stability and flow, implemented through partial symmetry breaking in recurrent connectivity. Circulant structure defines the internal representational manifold and antisymmetry determines how that manifold is traversed. Flow is therefore a functional requirement for hippocampal computation, enabling internal models to be dynamically explore rather than statically recalled. More broadly, this work suggest that CA3-like inductive bias does not prescribe a specific computation, but instead constrains learning toward a family of dynamical regimes well suited for sequential memory tasks—a tilted Mexican hat. Structured

connectivity restricts the search space of learning, while task demands select a particular operating point within that space. The convergence of mixed-symmetry structure across random initialization, structured bias, and biologically plausible plasticity argues that this organization is not a task artifact. Although the precise symmetry level, propagation timescale, and antisymmetric gain depend on task demands, the organizing principle of dominant symmetry combined with weak, derivative-like antisymmetry appears to be general.

5. Conclusion

Several limitations point to important directions for future work. We focused on a single prediction task and abstracted away cell-type diversity, neuromodulation, and short-term synaptic dynamics, all of which are known to influence hippocampal computation. Whether the same mixed-symmetry regime generalizes to non-spatial memories, branching sequences, or goal-dependent dynamics remains an open question. Extending this idea to multi-area models incorporating CA1, entorhinal cortex, and state-dependent modulation will be essential for linking circuit dynamics more directly to behavior.

The framework developed here yields clear experimental predictions. Selective perturbations of directional coupling within CA3 should impair replay speed or directionality without abolishing pattern completion, while state-dependent modulation of CA3 inputs should bias internal flow while preserving the underlying attractor geometry. Advances in large-scale electrophysiology and imaging now make it possible to directly test whether hippocampal population dynamics exhibit the predicted weak rotational components associated with partial symmetry breaking.

In conclusion, these results support a view of hippocampal CA3 as a flow-stabilized recurrent circuit in which memory, replay, and prediction arise from the interaction between inductive biases and experience-dependent learning. Partial symmetry breaking provides a principled mechanism for balancing stability and motion, offering a unifying circuit-level account of hippocampal sequence generation across behavioral states.

Acknowledgements. This research was supported by an NIH Director's Pioneer Award (DP1NS149613) and by a grant from the Office of Naval Research (N00014-23-1-2069).

References

1. O'Keefe J, Dostrovsky J. 1971 The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research* **34**, 171–175. (10.1016/0006-8993(71)90358-1)
2. Chen Y, Zhang H, Cameron M, Sejnowski T. 2024 Predictive Sequence Learning in the Hippocampal Formation. *Neuron* **112**. (https://doi.org/10.1016/j.neuron.2024.05.024)
3. Buzsáki G, Tingley D. 2018 Space and Time: The Hippocampus as a Sequence Generator. *Trends Cogn Sci* **22**, 853–869. (10.1016/j.tics.2018.07.006)
4. Treves A, Rolls ET. 1994 Computational analysis of the role of the hippocampus in memory. *Hippocampus* **4**, 374–391. (10.1002/hipo.450040319)
5. McNaughton BL, Morris RGM. 1987 Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci.* **10**, 408–415. (10.1016/0166-2236(87)90011-7)
6. Skaggs WE, McNaughton BL. 1992 Computational approaches to hippocampal function. *Curr. Opin. Neurobiol.* **2**, 209–211. (10.1016/0959-4388(92)90014-C)
7. Colgin LL, Leutgeb S, Jezek K, Leutgeb JK, Moser EI, McNaughton BL, Moser MB. 2010 Attractor-Map Versus Autoassociation Based Attractor Dynamics in the Hippocampal Network. *Journal of Neurophysiology* **104**, 35–50. (10.1152/jn.00202.2010)
8. Sheintuch L, Geva N, Deitch D, Rubin A, Ziv Y. 2023 Organization of hippocampal CA3 into correlated cell assemblies supports a stable spatial code. *Cell Reports* **42**, 112119. (10.1016/j.celrep.2023.112119)
9. Jezek K, Henriksen EJ, Treves A, Moser EI, Moser MB. 2011 Theta-paced flickering between place-cell maps in the hippocampus. *Nature* **478**, 246–249. (10.1038/nature10439)
10. Knierim JJ, Neunuebel JP. 2016 Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics. *Neurobiology of Learning and Memory* **129**, 38–49. (10.1016/j.nlm.2015.10.008)
11. O'Keefe J, Recce ML. 1993 Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317–330. (10.1002/hipo.450030307)
12. Skaggs WE, McNaughton BL, Wilson MA, Barnes CA. 1996 Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149–172. (10.1002/(SICI)1098-1063(1996)6:2<149::AID-HIPO6>3.0.CO;2-K)
13. Tsodyks MV, Skaggs WE, Sejnowski TJ, McNaughton BL. 1996 Population Dynamics and Theta Rhythm Phase Precession of Hippocampal Place Cell Firing: a Spiking Neuron Model. *Hippocampus* **6**, 271–280. (10.1002/(SICI)1098-1063(1996)6:3<271::AID-HIPO5>3.0.CO;2-Q)
14. Lisman JE. 1999 Relating Hippocampal Circuitry to Function: Recall of Memory Sequences by Reciprocal Dentate–CA3 Interactions. *Neuron* **22**, 233–242. (10.1016/S0896-6273(00)81085-5)
15. Johnson A, Redish AD. 2007 Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J Neurosci* **27**, 12176–12189. (10.1523/JNEUROSCI.3761-07.2007)
16. Manns JR, Howard MW, Eichenbaum H. 2007 Gradual changes in hippocampal activity support remembering the order of events. *Neuron* **56**, 530–540. (10.1016/j.neuron.2007.08.017)
17. Wilson MA, McNaughton BL. 1994 Reactivation of Hippocampal Ensemble Memories During Sleep. *Science* **265**, 676–679. (10.1126/science.8036517)
18. Pfeiffer BE, Foster DJ. 2013 Hippocampal place cell sequences depict future paths to remembered goals. *Nature* **497**, 74–79. (10.1038/nature12112)

19. Lubenov EV, Siapas AG. 2009 Hippocampal theta oscillations are travelling waves. *Nature* **459**, 534–539. (10.1038/nature08010)
20. Abbott LF, Blum KI. 1996 Functional Significance of Long-Term Potentiation for Sequence Learning and Prediction. *Cereb Cortex* **6**, 406–416. (10.1093/cercor/6.3.406)
21. Samsonovich A, McNaughton BL. 1997 Path Integration and Cognitive Mapping in a Continuous Attractor Neural Network Model. *J Neurosci* **17**, 5900–5920. (10.1523/JNEUROSCI.17-15-05900.1997)
22. Buzsáki G, Moser EI. 2013 Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nat Neurosci* **16**, 130–138. (10.1038/nn.3304)
23. Chen Z, Gomperts SN, Yamamoto J, Wilson MA. 2013 Neural Representation of Spatial Topology in the Rodent Hippocampus. *Neural computation* **26**, 1. (10.1162/NECO_a_00538)
24. Mehta MR, Lee AK, Wilson MA. 2002 Role of experience and oscillations in transforming a rate code into a temporal code. *Nature* **417**, 741–746. (10.1038/nature00807)
25. Carr MF, Jadhav SP, Frank LM. 2011 Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature Neuroscience* **14**, 147–153. (10.1038/nn.2732)
26. Davidson TJ, Kloosterman F, Wilson MA. 2009 Hippocampal Replay of Extended Experience. *Neuron* **63**, 497–507. (10.1016/j.neuron.2009.07.027)
27. Gupta AS, van der Meer MA, Touretzky DS, Redish AD. 2010 Hippocampal Replay Is Not a Simple Function of Experience. *Neuron* **65**, 695–705. (10.1016/j.neuron.2010.01.034)
28. Ólafsdóttir HF, Bush D, Barry C. 2018 The Role of Hippocampal Replay in Memory and Planning. *Current Biology* **28**, R37–R50. (10.1016/j.cub.2017.10.073)
29. Bendor D, Wilson MA. 2012 Biasing the content of hippocampal replay during sleep. *Nature Neuroscience* **15**, 1439–1444. (10.1038/nn.3203)
30. Mattar MG, Daw ND. 2018 Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience* **21**, 1609–1617. (10.1038/s41593-018-0232-z)
31. O'Keefe J, Burgess N. 2005 Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. *Hippocampus* **15**, 853–866. (10.1002/hipo.20115)
32. Amit DJ, Tsodyks MV. 1991 Quantitative study of attractor neural network retrieving at low spike rates:: I. substrate—spikes, rates and neuronal gain. *Network: Computation in Neural Systems* **2**, 259–273. (10.1088/0954-898X_2_3_003)
33. Amit D, Tsodyks M. 1992 Effective neurons and attractor neural networks in cortical environment. *Network* **3**, 121–137. (10.1088/0954-898X/3/2/003)
34. Amit DJ, Brunel N, Tsodyks MV. 1994 Correlations of cortical Hebbian reverberations: theory versus experiment. *J Neurosci* **14**, 6435–6445. (10.1523/JNEUROSCI.14-11-06435.1994)
35. Kim R, Li Y, Sejnowski TJ. 2019 Simple framework for constructing functional spiking recurrent neural networks. *Proceedings of the National Academy of Sciences* **116**, 22811–22820. (10.1073/pnas.1905926116)
36. Hopfield JJ. 1982 Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A* **79**, 2554–2558. (10.1073/pnas.79.8.2554)
37. Amari Si. 1977 Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* **27**, 77–87. (10.1007/BF00337259)
38. Brunel N. 2016 Is cortical connectivity optimized for storing information?. *Nature Neuroscience* **19**, 749–755. (10.1038/nn.4286)
39. Griniasty M, Tsodyks MV, Amit DJ. 1993 Conversion of Temporal Correlations Between Stimuli to Spatial Correlations Between Attractors. *Neural Computation* **5**, 1–17. (10.1162/neco.1993.5.1.1)
40. Cugliandolo LF, Tsodyks MV. 1994 Capacity of networks with correlated attractors. *J. Phys. A: Math. Gen.* **27**, 741. (10.1088/0305-4470/27/3/018)
41. Seung H. 1996 How the brain keeps the eyes still. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13339–13344. (10.1073/pnas.93.23.13339)
42. Eliasmith C. 2005 A Unified Approach to Building and Controlling Spiking Attractor Networks. *Neural Computation* **17**, 1276–1314. (10.1162/0899766053630332)
43. Kleinfeld D. 1986 Sequential state generation by model neural networks.. *Proceedings of the National Academy of Sciences* **83**, 9469–9473. (10.1073/pnas.83.24.9469)
44. Sompolinsky H, Kanter I. 1986 Temporal Association in Asymmetric Neural Networks. *Phys. Rev. Lett.* **57**, 2861–2864. (10.1103/PhysRevLett.57.2861)
45. Levy WB. 1996 A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus* **6**, 579–590. (10.1002/(sici)1098-1063(1996)6:6<579::aid-hipo3>3.0.co;2-c)
46. Jensen O, Lisman JE. 1996 Hippocampal CA3 region predicts memory sequences: accounting for the phase precession of place cells.. *Learning amp; Memory* **3**, 279–287. (10.1101/lm.3.2-3.279)
47. Hasselmo ME, Eichenbaum H. 2005 Hippocampal mechanisms for the context-dependent retrieval of episodes. *Neural Networks* **18**, 1172–1190. (10.1016/j.neunet.2005.08.007)
48. Widloski J, Foster DJ. 2025 Replay without sharp wave ripples in a spatial memory task. *Nature Communications* **16**. (10.1038/s41467-025-65181-5)
49. Ji Z, Chu T, Wu S, Burgess N. 2025 A systems model of alternating theta sweeps via firing rate adaptation. *Current Biology* **35**, 709–722.e5. (10.1016/j.cub.2024.08.059)
50. Chandra S, Sharma S, Chaudhuri R, Fiete I. 2025 Episodic and associative memory from spatial scaffolds in the hippocampus. *Nature* **638**, 739–751. (10.1038/s41586-024-08392-y)
51. Hasselmo ME. 2009 A model of episodic memory: Mental time travel along encoded trajectories using grid cells. *Neurobiology of Learning and Memory* **92**, 559–573. (10.1016/j.nlm.2009.07.005)
52. Agmon H, Burak Y. 2020 A theory of joint attractor dynamics in the hippocampus and the entorhinal cortex accounts for artificial remapping and grid cell field-to-field variability. *eLife* **9**. (10.7554/elife.56894)
53. Zhang K. 1996 Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.* **16**, 2112–2126. (10.1523/JNEUROSCI.16-06-02112.1996)

54. Rajan K, Harvey CD, Tank DW. 2016 Recurrent Network models of sequence generation and memory. *Neuron* **90**, 128–142. (10.1016/j.neuron.2016.02.009)
55. Keller TA, Muller L, Sejnowski T, Welling M. 2024 Traveling Waves Encode the Recent Past and Enhance Sequence Learning. ICLR. arXiv:2309.08045 [cs] (10.48550/arXiv.2309.08045)
56. Mohan UR, Zhang H, Ermentrout B, Jacobs J. 2024 The direction of theta and alpha travelling waves modulates human memory processing. (10.1038/s41562-024-01838-3)
57. Sompolinsky H, Crisanti A, Sommers HJ. 1988 Chaos in Random Neural Networks. *Phys. Rev. Lett.* **61**, 259–262. (10.1103/PhysRevLett.61.259)
58. Bush D, Philippides A, Husbands P, O'Shea M. 2010 Dual Coding with STDP in a Spiking Recurrent Neural Network Model of the Hippocampus. *PLoS Comput Biol* **6**, e1000839. (10.1371/journal.pcbi.1000839)
59. Romani S, Tsodyks M. 2015 Short-term plasticity based network model of place cells dynamics. *Hippocampus* **25**, 94–105. (10.1002/hipo.22355)
60. Roxin A, Brunel N, Hansel D. 2005 Role of Delays in Shaping Spatiotemporal Dynamics of Neuronal Activity in Large Networks. *Phys. Rev. Lett.* **94**, 238103. (10.1103/PhysRevLett.94.238103)
61. Pouget A, Sejnowski TJ. 1995 pp. 368–371. In *Dynamic Remapping*, pp. 368–371. Cambridge, MA: MIT Press 2 edition.
62. Kurikawa T, Kaneko K. 2021 Multiple-timescale neural networks: Generation of history-dependent sequences and inference through autonomous bifurcations. *Front. Comput. Neurosci.* **15**, 743537. (10.3389/fncom.2021.743537)
63. Betteti S, Baggio G, Bullo F, Zampieri S. 2025 Input-driven dynamics for robust memory retrieval in Hopfield networks. *Science Advances* **11**, eadu6991. (10.1126/sciadv.adu6991)
64. Karuvally A, Sejnowski T, Siegelmann HT. 2023 General Sequential Episodic Memory Model. In Krause A, Brunskill E, Cho K, Engelhardt B, Sabato S, Scarlett J, editors, *Proceedings of the 40th International Conference on Machine Learning* vol. 202*Proceedings of Machine Learning Research* pp. 15900–15910. PMLR.
65. Karuvally A, Lertsaroj P, Sejnowski T, Siegelmann HT. 2025 Exponential Dynamic Energy Network for High Capacity Sequence Memory. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.
66. Patel J, Fujisawa S, Berényi A, Royer S, Buzsáki G. 2012 Traveling theta waves along the entire septotemporal axis of the hippocampus. *Neuron* **75**, 410–417. (10.1016/j.neuron.2012.07.015)
67. Markram H, Lübke J, Frotscher M, Sakmann B. 1997 Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **275**, 213–215. (10.1126/science.275.5297.213)
68. Bi GQ, Poo MM. 1998 Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J Neurosci* **18**, 10464–10472. (10.1523/JNEUROSCI.18-24-10464.1998)
69. Mishra RK, Kim S, Guzman SJ, Jonas P. 2016 Symmetric spike timing-dependent plasticity at CA3–CA3 synapses optimizes storage and recall in autoassociative networks. *Nat Commun* **7**, 11552. (10.1038/ncomms11552)
70. Gillett M, Brunel N. 2024 Dynamic control of sequential retrieval speed in networks with heterogeneous learning rules. *eLife* **12**, RP88805. (10.7554/eLife.88805)
71. Sammons RP, Masserini S, Moreno Velasquez L, Metodieva VD, Cano G, Sannio A, Orlando M, Maier N, Kempter R, Schmitz D. 2025 Sub-type specific connectivity between CA3 pyramidal neurons may underlie their sequential activation during sharp waves. *eLife* **13**, RP98653. (10.7554/eLife.98653)
72. Watson JF, Vargas-Barroso V, Jonas P. 2025 Cell-specific wiring routes information flow through hippocampal CA3. *Cell Rep.* **44**, 116080. (10.1016/j.celrep.2025.116080)
73. Layous R, Tamás B, Mike A, Sipos E, Arszovszki A, Brunner J, Szatai Á, Yaseen F, Andrási T, Szabadics J. 2025 Optical recordings of unitary synaptic connections reveal high and random local connectivity between CA3 pyramidal cells. *J. Neurosci.* **45**, e0102252025. (10.1523/JNEUROSCI.0102-25.2025)
74. Chenkov N, Sprekeler H, Kempter R. 2017 Memory replay in balanced recurrent networks. *PLoS Comput Biol* **13**, e1005359. (10.1371/journal.pcbi.1005359)
75. Laje R, Buonomano DV. 2013 Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nat Neurosci* **16**, 925–933. (10.1038/nn.3405)
76. Buzsáki G. 2010 Neural syntax: cell assemblies, synapsembles and readers. *Neuron* **68**, 362–385. (10.1016/j.neuron.2010.09.023)
77. Watson JF, Vargas-Barroso V, Morse-Mora RJ, Navas-Olive A, Tavakoli MR, Danzl JG, Tomschik M, Rössler K, Jonas P. 2025 Human hippocampal CA3 uses specific functional connectivity rules for efficient associative memory. *Cell* **188**, 501–514.e18. (10.1016/j.cell.2024.11.022)
78. Williams RJ, Zipser D. 1989 A learning algorithm for continually running fully recurrent neural networks. *Neural Comput.* **1**, 270–280. (10.1162/neco.1989.1.2.270)
79. Ben-Yishai R, Bar-Or RL, Sompolinsky H. 1995 Theory of orientation tuning in visual cortex.. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3844–3848. (10.1073/pnas.92.9.3844)
80. Liu YH, Baratin A, Cornford J, Mihalas S, Shea-Brown E, Lajoie G. 2024 How connectivity structure shapes rich and lazy learning in neural circuits. arXiv:2310.08513 [cs] (10.48550/arXiv.2310.08513)
81. Bordelon B, Cotler J, Pehlevan C, Zavatone-Veth JA. 2025 Dynamically Learning to Integrate in Recurrent Neural Networks. arXiv:2503.18754 [q-bio] (10.48550/arXiv.2503.18754)
82. Keller TA. 2025 Flow equivariant recurrent neural networks. In *Proceedings of the 39th International Conference on Neural Information Processing Systems*.
83. Khona M, Fiete IR. 2022 Attractor and integrator networks in the brain. *Nat Rev Neurosci* **23**, 744–766. (10.1038/s41583-022-00642-0)
84. Sussillo D, Abbott LF. 2009 Generating Coherent Patterns of Activity from Chaotic Neural Networks. *Neuron* **63**, 544–557. (10.1016/j.neuron.2009.07.018)
85. Dubreuil A, Valente A, Beiran M, Mastrogiuseppe F, Ostojic S. 2022 The role of population structure in computations through neural dynamics. *Nat Neurosci* **25**, 783–794. (10.1038/s41593-022-01088-4)
86. Vandael D, Okamoto Y, Jonas P. 2021 Transsynaptic modulation of presynaptic short-term plasticity in hippocampal mossy fiber synapses. *Nat Commun* **12**, 2912. (10.1038/s41467-021-23153-5)

87. Song HF, Yang GR, Wang XJ. 2016 Training Excitatory-Inhibitory Recurrent Neural Networks for Cognitive Tasks: A Simple and Flexible Framework. *PLOS Computational Biology* **12**, e1004792. (10.1371/journal.pcbi.1004792)
88. Barreiro AK, Kutz JN, Shlizerman E. 2017 Symmetries Constrain Dynamics in a Family of Balanced Neural Networks. *The Journal of Mathematical Neuroscience* **7**. (10.1186/s13408-017-0052-6)
89. Sourmpis C, Petersen CC, Gerstner W, Bellec G. 2026 Biologically informed cortical models predict optogenetic perturbations. *eLife* **14**. (10.7554/elife.106827)
90. Masoliver M, Davidsen J, Nicola W. 2022 Embedded chimera states in recurrent neural networks. *Communications Physics* **5**. (10.1038/s42005-022-00984-2)
91. Vogels TP, Sprekeler H, Zenke F, Clopath C, Gerstner W. 2011 Inhibitory Plasticity Balances Excitation and Inhibition in Sensory Pathways and Memory Networks. *Science* **334**, 1569–1573. (10.1126/science.1211095)
92. Guzman SJ, Schlögl A, Frotscher M, Jonas P. 2016 Synaptic mechanisms of pattern completion in the hippocampal CA3 network. *Science* **353**, 1117–1123. (10.1126/science.aaf1836)
93. Le Duigou C, Simonnet J, Teleñczuk MT, Fricker D, Miles R. 2014 Recurrent synapses and circuits in the CA3 region of the hippocampus: an associative network. *Frontiers in Cellular Neuroscience* **7**. (10.3389/fncel.2013.00262)
94. Miles R, Wong RK. 1986 Excitatory synaptic interactions between CA3 neurones in the guinea-pig hippocampus.. *The Journal of Physiology* **373**, 397–418. (10.1113/jphysiol.1986.sp016055)
95. Lee JS, Briguglio JJ, Cohen JD, Romani S, Lee AK. 2020 The Statistical Structure of the Hippocampal Code for Space as a Function of Time, Context, and Value. *Cell* **183**, 620–635.e22. (10.1016/j.cell.2020.09.024)
96. Agarwal G, Stevenson IH, Berényi A, Mizuseki K, Buzsáki G, Sommer FT. 2014 Spatially Distributed Local Fields in the Hippocampus Encode Rat Position. *Science* **344**, 626–630. (10.1126/science.1250444)
97. Mizuseki K, Royer S, Diba K, Buzsáki G. 2012 Activity dynamics and behavioral correlates of CA3 and CA1 hippocampal pyramidal neurons. *Hippocampus* **22**, 1659–1680. (10.1002/hipo.22002)
98. Leutgeb S, Leutgeb JK, Treves A, Moser MB, Moser EI. 2004 Distinct Ensemble Codes in Hippocampal Areas CA3 and CA1. *Science* **305**, 1295–1298. (10.1126/science.1100265)
99. Foster DJ, Wilson MA. 2006 Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440**, 680–683. Publisher: Nature Publishing Group (10.1038/nature04587)
100. Peng Y, Bjelde A, Aceituno PV, Mittermaier FX, Planert H, Grosser S, Onken J, Faust K, Kalbhenn T, Simon M, Radbruch H, Fidzinski P, Schmitz D, Alle H, Holtkamp M, Vida I, Grewe BF, Geiger JRP. 2024 Directed and acyclic synaptic connectivity in the human layer 2-3 cortical microcircuit. *Science* **384**, 338–343. (10.1126/science.adg8828)
101. Oishi N, Nomoto M, Ohkawa N, Saitoh Y, Sano Y, Tsujimura S, Nishizono H, Matsuo M, Muramatsu Si, Inokuchi K. 2019 Artificial association of memory events by optogenetic stimulation of hippocampal CA3 cell ensembles. *Mol Brain* **12**, 2. (10.1186/s13041-018-0424-1)
102. Quentin R, King JR, Sallard E, Fishman N, Thompson R, Buch ER, Cohen LG. 2019 Differential brain mechanisms of selection and maintenance of information during working memory. *J. Neurosci.* **39**, 3728–3740. (10.1523/JNEUROSCI.2764-18.2019)
103. Durbin J, Watson GS. 1971 Testing for serial correlation in least squares regression. III. *Biometrika* **58**, 1. (10.2307/2334313)
104. Batschelet E. 1983 Circular statistics in biology. *J. R. Stat. Soc. Ser. A* **146**, 91. (10.2307/2981498)

6. Supplemental Material

(a) Pretraining as evidence for structural inductive bias

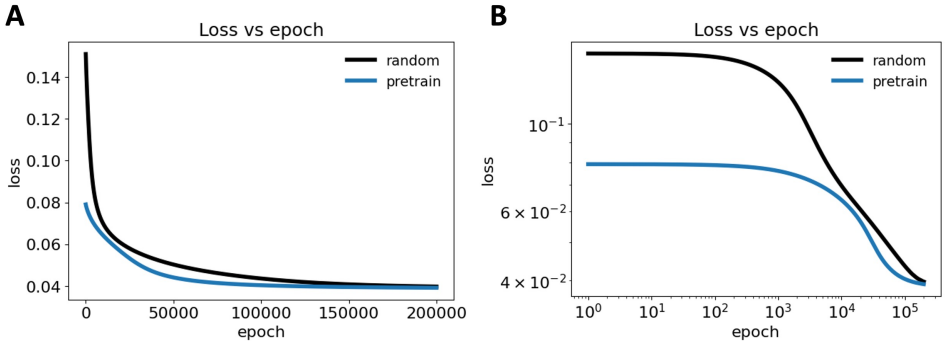

Figure S1: **Effect of pretraining on learning dynamics. (A)** Training loss as a function of epoch shown on linear axes. Networks initialized with pretrained recurrent weights (blue), obtained from the final weights of the randomly initialized network, begin at a lower loss and exhibit faster early optimization compared to networks trained from random initialization (black). **(B)** Same data plotted on log-log axes, highlighting differences in early and intermediate training regimes. Pretrained networks avoid the prolonged high-loss plateau observed under random initialization and transition earlier into the rapid loss-decrease regime. In both cases, training converges to a similar final loss.

Pretraining the recurrent connectivity using the final weights obtained from a randomly initialized network significantly alters the early learning dynamics without changing the eventual performance ceiling (Fig. S1). As shown in both linear (Fig. S1A) and log-log (Fig. S1B) axes, pretrained networks start at a markedly lower loss and bypass the extended high-loss plateau characteristic of random initialization. This leads to faster progress through early and intermediate training epochs. Despite these differences in learning trajectory, both conditions converge to nearly identical final loss values, indicating that pretraining primarily improves optimization efficiency rather than the final representational capacity of the model. These results suggest that structure acquired during prior training places the network in a favorable regime of parameter space, facilitating faster learning while reserving the same asymptotic solution.

(b) Model hyperparameters

Table S1: Model hyperparameters

hyperparameter	value
number of units	100
number of epochs	100,000
sequence length	100
learning rate	0.01

(c) Identity and cyclic shift initializations: isolating persistence, directionality, and their limits in CA3-like circuits

To further dissect which structural features of CA3 recurrent connectivity are necessary for hippocampal-like replay and prediction, we examined two extreme and deliberately simplified structures: identity and cyclic shift connectivity. These architectures isolate the complementary computational primitives of temporal persistence and directional propagation (symmetric and antisymmetric aspects) that are each present in CA3, although not in isolation.

The identity matrix represents a maximally symmetric, self-coupled network in which each unit predominantly feeds back onto itself. We include this case as a conceptual control that isolates temporal persistence without lateral interactions. Such a network can maintain activity locally and support short-horizon, input-driven prediction. However, it fundamentally fails to capture CA3's role as an associative recurrent circuit. Identity-initialized networks reliably learned one-step prediction when driven by external input but failed at autonomous replay. The resulting weight profiles remained identity-dominated with only weak lateral structure, which was insufficient to sustain a self-organized sequence when driven by noise. Thus, persistence alone is not sufficient for sequential dynamics or replay. This reinforces the idea that CA3 replay requires not just stability, but structured interactions that organize population activity along a low-dimensional manifold.

In contrast, cyclic shift connectivity (Fig. **??**) imposes a purely directional recurrent structure, in which each neuron excites its success along a ring. This architecture hardwires a translation operator into the recurrent matrix and naturally produces a

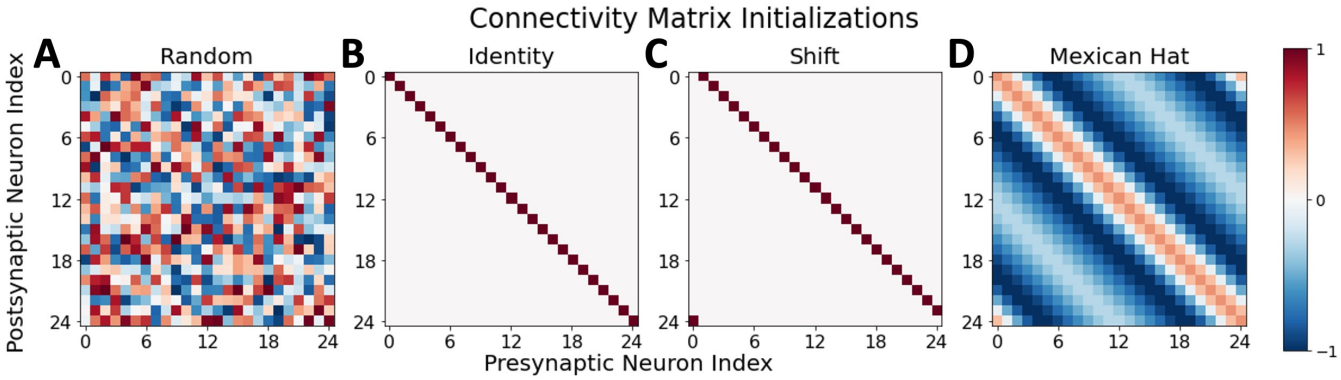

Figure S2: **Connectivity matrix initialization families** Visualization of the four recurrent families used in this study: **(A)** random connectivity, with weights sampled from a zero-mean distribution (PyTorch baseline). **(B)** identity connectivity, in which each neuron connects only to itself. **(C)** cyclic shift connectivity, where each neuron projects to its immediate successor on a ring (with wrap-around), implementing a discrete translation operator. **(D)** Mexican-hat connectivity, constructed as a difference-of-Gaussians kernel on a ring, producing local excitation and broader surround inhibition. Identity and cyclic shift connectivity represent limiting cases of purely symmetric (self-coupled) and purely directional (translation-only) structure, respectively, while Mexican-hat connectivity combines structured excitation and inhibition to support attractor dynamics.

sequential propagation of activity. As such, it captures the CA3 function of directional progression through a learned sequence and is consistent with experimental and theoretical work implicating antisymmetric or non-reciprocal connectivity arising from temporally antisymmetric plasticity such as STDP. Indeed, antisymmetric components in CA3 may reflect experience-dependent strengthening of forward transitions during repeated trajectories, theta sequences, or sharp-wave ripple replay.

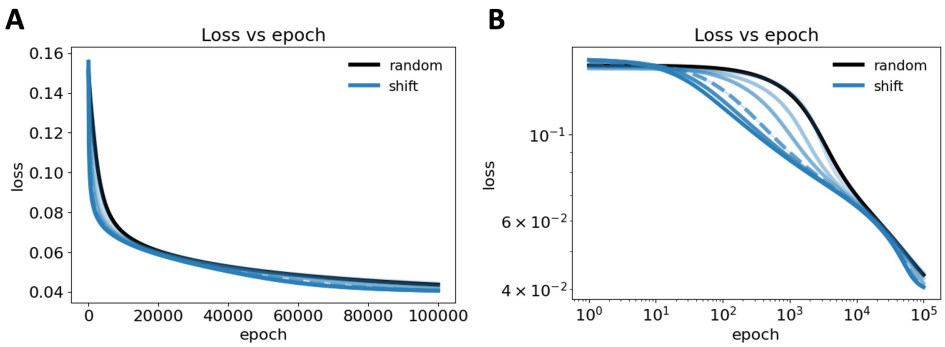

Figure S3: **Training loss for cyclic shift-initialized recurrent networks across symmetry levels.** Training loss as a function of epoch for cyclic shift initializations (blue) spanning initial symmetry ratios of $\alpha_0 \in [0, 1]$ compared to random initialization (black). Light blue is the most antisymmetric and the symmetry ratio increases as the blue line darkens. The dashed blue curve is $\alpha_0 = 0.60$, the beginning of the regime where replay breaks down despite comparable training loss. **(A)** linear scale and **(B)** log-log scale. Overall, cyclic shift-initialized networks exhibit similar final and initial losses to random connectivity, while early training dynamics are altered. Stable replay here is not guaranteed.

However, pure cyclic shift connectivity lacks the bidirectional, stabilizing interactions required for attractor-like memory retrieval. Our results make this limitation explicit. While cyclic shift-initialized networks were able to outperform identity networks on both replay and prediction, their ability to support replay depended critically on the initial symmetry-antisymmetry balance. As the initial symmetry ratio α_0 increased, the structure became more similar to the identity matrix, with a complete sharp breakdown in replay behavior at $\alpha_0 = 0.60$ indicated by ring-decode R^2 collapse, residual temporal coherence deterioration, and polar trajectories which degenerated into fixed or noisy states (Fig. S4-S5) Indeed, at $\alpha_0 = 0$, the recurrent matrix is a pure cyclic shift matrix, but at $\alpha_0 = 1$, the matrix has the same limitations as the identity matrix. Notably, prediction performance remained largely intact across this transition, indicating that locally accurate state updates can still be learned even when the global autonomous dynamics fail.

This highlights a key computational distinction. Prediction can be supported by a locally correct transition operator under external drive, whereas replay requires a globally consistent dynamical system with both a stable manifold and a directed flow along it. Pure cyclic shift connectivity provides direction without sufficient stabilization; identity provides stabilization

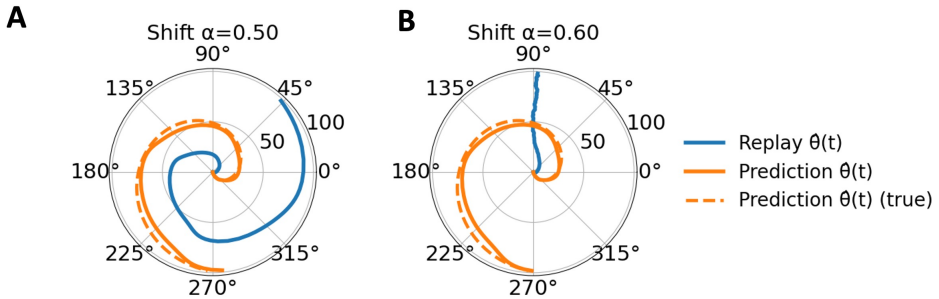

Figure S4: **Breakdown of replay for cyclic shift-initialized networks at higher symmetry**. Polar trajectories of learned replay (blue) and prediction (orange; dashed indicated ground truth) for cyclic shift-initialized recurrent connectivity with **(A)** $\alpha_0 = 0.50$ and **(B)** $\alpha_0 = 0.60$. At $\alpha_0 = 0.50$, replay follows a coherent progression around the ring, whereas increasing symmetry to $\alpha_0 = 0.60$ leads to a collapse of replay dynamics despite preserved prediction performance, indicating a sharp transition in the ability of cyclic shift connectivity to support stable autonomous sequence generation.

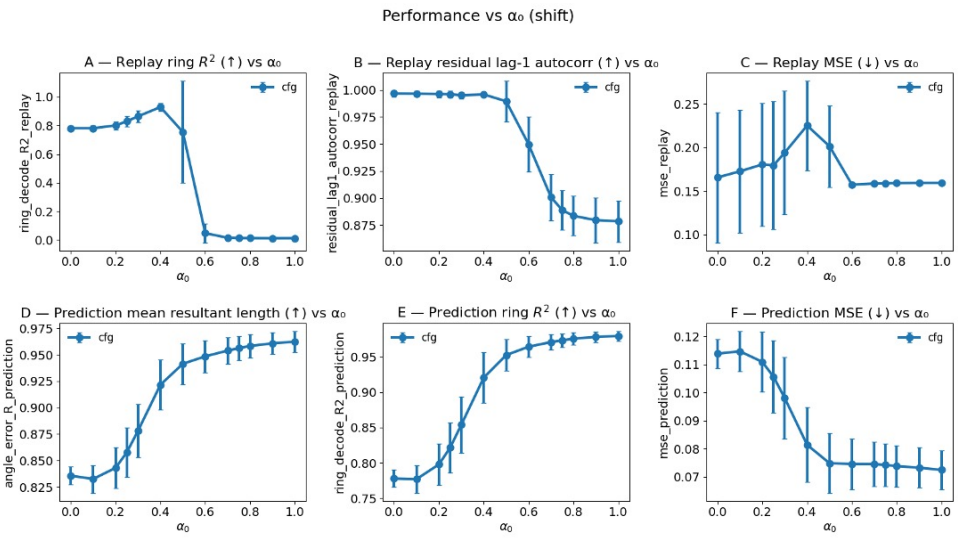

Figure S5: **Performance as a function of initial symmetry for cyclic shift connectivity.** Replay and prediction performance metrics plotted as a function of the initial symmetry ratio α_0 for cyclic shift-initialized recurrent connectivity. Top row: replay metrics including ring-decode R^2 **(A)**, residual lag-1 autocorrelation **(B)**, and replay MSE **(C)**. Bottom row: prediction metrics including mean resultant length of angular error **(D)**, ring-decode R^2 **(E)**, and prediction MSE **(F)**. Replay performance exhibits a sharp breakdown at $\alpha_0 = 0.60$, marked by the dramatic drop in ring-decode R^2 and reduced temporal coherence despite increased prediction performance. This difference indicates that increasing symmetry in cyclic shift connectivity impairs autonomous replay dynamics while leaving input-driven prediction intact.

without direction. Random connectivity can, through learning, self-organize both to some degree but does so less intuitively than biologically inspired structures.

Across structured initializations, training consistently drove the recurrent connectivity toward an intermediate symmetry regime, quantified by convergence of $\alpha(t)$ (Fig. S6). Fully symmetric networks (i.e. identity) rapidly gained antisymmetric structure. Cyclic shift networks, spanning fully antisymmetric to symmetric initial conditions, converged toward a mid-range α value. Of note, the intermediate range here is lower than that of Mexican-hat initialization. This is likely connected to the breakdown in replay behavior observed at $\alpha_0 = 0.60$. This convergence is again not an artifact of initialization, but instead an emergent, task-selected property, reinforcing the conclusion that hippocampal-like computation requires a balance between symmetric stabilization and antisymmetric flow.

A tilted Mexican-hat connectivity obtains this balance naturally. By providing a center-surround attractor scaffold from the outset, Mexican-hat initialization supports both stable memory representations and controlled propagation, yielding superior replay fidelity, tighter alignment between replay and prediction than either cyclic shift or identity connectivity. In this context, the cyclic shift and identity initializations serve as informative boundary cases that reveal that neither directionality nor persistence alone is sufficient. Instead, CA3-like replay emerges most robustly when directional biases are embedded with a stabilizing recurrent architecture. This is a regime that learning reliably discovers, but that structured inductive bias makes easier to reach and clearer to interpret.

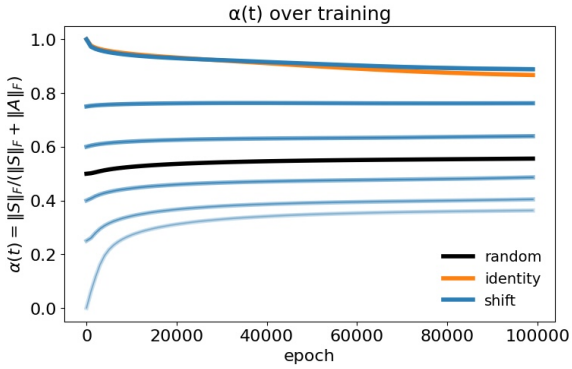

Figure S6: **Training drives recurrent connectivity toward an intermediate symmetry regime.** Evolution of the symmetry index $\alpha(t) = ||S||_F/(||S||_F + ||A||_F)$ during training for networks initialized with random (black), identity (orange), and cyclic shift (blue; multiple initial α_0) recurrent connectivity. The initial α value, α_0, is given by the value at epoch 0. Across initial conditions, learning consistently drives $\alpha(t)$ toward an intermediate range, indicating convergence to a mixed symmetric-antisymmetric recurrent regime.

Together, these results argue that CA3 is best understood not as a pure attractor (identity) network or a feedforward sequence generator (pure cyclic shift), but as a flow-stabilized recurrent system in which symmetric and antisymmetric interactions are jointly essential. Identity and cyclic shift initializations clarify the computational roles of these components, while Mexican-hat connectivity demonstrates how their integration supports replay and prediction in a single recurrent circuit.

(d) Effective Jacobian and gain statistics

In an RNN with $\tanh$ nonlinearity, stability is determined by the instantaneous Jacobian

$$J_t = \mathrm{diag}(1 - h_t^2)W_{hh} \tag{6.1}$$

where h_t denotes the hidden state.

Because the network operates in a driven regime without a single fixed point, the relevant operating regime is defined by trajectories during replay and prediction. We therefore quantified the effective gain of the $\tanh$ nonlinearity along these trajectories. The $\tanh$ derivative acts as a state-dependent gain that rescales the recurrent matrix, determining the effective dynamics.

Across runs with random initialization, the mean derivative during replay was 0.31 ± 0.005 (median 0.14), with approximately 43% of activations having derivative less than 0.1, indicating substantial saturation (see Fig. S7 for full distribution). Similar values were observed during prediction (mean ≈ 0.27). Networks initialized with Mexican-hat connectivity exhibited slightly lower gains (mean 0.23-0.27), consistent with a more saturated operating regime.

These results imply that the effective Jacobian is a state-dependent, diagonally scaled version of W_{hh}, which contracts the spectrum relative to the raw recurrent matrix. Consequently, eigenvalues of W_{hh} exceeding unity do not imply instability of the nonlinear system. Instead, the eigenspectrum of W_{hh} should be interpreted as reflecting the dominant recurrent interaction structure, with the nonlinear gain determining the effective dynamical regime.

Table S2: Gain analysis

Condition	Mean gain	Median gain	Percent < 0.1
Random replay	0.305 ± 0.005	.0142	43%
Random prediction	0.272 ± 0.002	0.116	47%
Mexican hat replay	0.268 ± 0.005	0.083	53%
Mexican hat prediction	0.228 ± 0.003	0.062	58%

The broad distribution of tanh gain indicates that the network operates in a heterogeneous dynamical regime rather than a globally linear one. Many units are partially saturated, which is consistent with recurrent stabilization, while a smaller subset remains in a higher-gain regime capable of supporting directional state transitions. This is qualitatively consistent with the connectivity decomposition into a dominant symmetric scaffold plus a weaker antisymmetric component.

(e) Metrics computed

(i) Ring decoding via cosine-sine regression

We evaluate decoding of a circular variable using linear (ridge) regression on cosine and sine targets, reporting performance via R^2 [102]. Decoding $\cos\theta$ and $\sin\theta$ avoids angular wrap-around discontinuities.

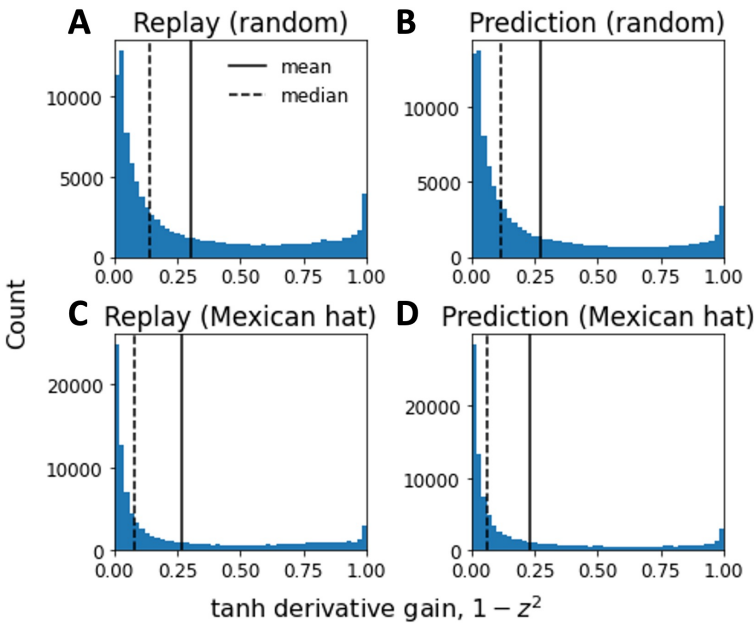

Figure S7: **Effective gain of** $\tanh$ **nonlinearity during replay and prediction** Distribution of the instantaneous derivative $1 - h^2$ across hidden units and time evaluated for replay and prediction trajectories for networks with random and Mexican-hat initialization. Vertical lines indicate mean (solid) and median (dashed). The contraction factor (0.25-0.3) suggests effective spectral radius is reduced proportionally. In all conditions, the network operates in a moderately saturated regime, with mean gain of 0.2-0.3 and a substantial fraction of units exhibiting low gain (<0.1). Because the Jacobian is driven by $J_t = \mathrm{diag}(1 - h_t^2)W_{hh}$, these results indicate that the effective dynamics correspond to a gain-scaled version of the recurrent matrix, contracting the spectrum relative to W_{hh}.

At each time step, the network produces an N-dimensional population activity vector interpreted as a ring code. Outputs are rectified and normalized to form a population distribution,

$$x_t = \mathrm{normalize}(y_t) = \frac{\max(y_t, 0)}{\sum_{i=1}^{N} \max(y_{t,i}, 0) + \epsilon} \tag{6.2}$$

Each unit i is assigned a preferred angle $\phi_i = 2\pi i/N$. Given the target ring distribution p_t, the target angle is computed via its first circular moments

$$C_t = \sum_{i=1}^{N} p_{t,i}\cos(\phi_i), \quad S_t = \sum_{i=1}^{N} p_{t,i}\sin(\phi_i), \quad \theta_t = \mathrm{atan2}(S_t, C_t) \tag{6.3}$$

and represented as $Y_t = [\cos\theta_t, \sin\theta_t]$

Stacking all time points yields $X \in \mathbb{R}^{T\times N}$ and $Y \in \mathbb{R}^{T\times 2}$. A ridge decoder is fit as

$$W = (X^\top X + \lambda I)^{-1} X^\top Y, \quad \hat{Y} = XW \tag{6.4}$$

Performance is quantified by computing R^2 separately for cosine and sine and averaging

$$\text{ring decode } R^2 = \frac{1}{2}\left(R^2_{\cos} + R^2_{\sin}\right) \tag{6.5}$$

This metric measures time-resolved fidelity to the trained target phase θ_t. It is not invariant to phase shifts, speed changes, or temporal offsets. For replay, it reflects spontaneous reproduction of the learned sequence under noise input. For prediction, it reflects accuracy in tracking the target trajectory given an initial drive. Values lie in $(-\infty, 1]$, with $R^2 = 1$ indicating a correctly positioned, stable activity bump at all time steps.

(ii) Lag-1 autocorrelation

Lag-1 autocorrelation quantifies temporal dependence in the magnitude of model residuals. For each time step, we compute the residual magnitude

$$r_t = ||\text{Target}_t - \text{Output}_t||_2 \tag{6.6}$$

Then demeaning the resulting time series

$$x_t = r_t - \text{mean}_t(r_t) \tag{6.7}$$

and compute the correlation between $(x_1, ..., x_{T-1})$ and $(x_2, ..., x_T)$ (standard lag-1 autocorrelation).

Values lie in $[-1, 1]$. Zero indicates temporally uncorrelated (white-noise-like) residuals, positive values indicate smooth temporal persistence of errors, and negative values indicate alternating structure. High lag-1 autocorrelation reflects slowly varying residual magnitudes consistent with coherent underlying dynamics [103].

(iii) Mean resultant length (angular error concentration)

This metric quantifies how concentrated the circular angular error $\Delta\theta(t)$ is over time between the true angle implied by the target ring distribution and the angle decoded from the model output [104]. It is also known as the mean resultant length, vector strength, or phase-locking value.

At each time step, target and model output activities are converted to angles, and the wrapped angular error $\Delta\theta_t$ is computed. Each error is represented as a unit vector

Let $\Delta\theta_{b,t}$ be the wrapped circular error between the predicted and ground truth angle. We can define unit vectors as

Then demean according to

$$u_t = (\cos\Delta\theta_t, \sin\Delta\theta_t) \tag{6.8}$$

which are averaged over time

$$\bar{u} = (\langle\cos\Delta\theta\rangle, \langle\sin\Delta\theta\rangle) \tag{6.9}$$

The mean resultant length is

$$R = ||\bar{u}|| = \sqrt{\langle\cos\Delta\theta\rangle^2 + \langle\sin\Delta\theta\rangle^2} \tag{6.10}$$

$R \in [0, 1]$, where $R = 1$ indicates highly consistent angular errors (strong concentration near a single phase, typically zero) and $R = 0$ indicates uniformly distributed, random angular errors.

(f) Symmetric STDP Analysis

The analytic form of the final weights is derived by convolving the Effective STDP Kernel (which includes the symmetry correction ϵ) with the Autocorrelation of the traveling Gaussian input. This is derived from the STDP learning kernel defined as

$$W_{ij} = \int K(s) \int u_i^{(t)}\, u_j^{(t-s)}\, \mathrm{d}t\, \mathrm{d}s + \epsilon \int K(s) \int u_j^{(t)}\, u_i^{(t-s)}\, \mathrm{d}t\, \mathrm{d}s \tag{6.11}$$

where

$$K(s) = \begin{cases} A_+\, e^{-s/\tau_+} & s \geq 0 \\ -A_-\, e^{s/\tau_-} & s < 0 \end{cases} \quad \text{and} \quad u_i^{(t)} = \mathcal{N}(a, b, \mu)[i] = a e^{-b\,(i-(\mu+t))^2} \tag{6.12}$$

First computing the product $u_i^{(t)} u_j^{(t-s)}$

$$u_i^{(t)}\, u_j^{(t-s)} = \mathcal{N}(a, b, \mu + t)[i]\, \mathcal{N}(a, b, \mu + t - s)[j] \tag{6.13}$$

$$u_i^{(t)}\, u_j^{(t-s)} = \mathcal{N}(a, b, \mu + t)[i]\, \mathcal{N}(a, b, i - j + \mu + t - s)[i] \tag{6.14}$$

$$u_i^{(t)}\, u_j^{(t-s)} = \mathcal{N}(a, b, \mu - i)[-t]\, \mathcal{N}(a, b, \mu - (j + s))[-t] \tag{6.15}$$

Using Lemma 1,

$$u_i^{(t)}\, u_j^{(t-s)} = \mathcal{N}\left(a^2\, e^{-\frac{b(s+j-i)^2}{2}}, 2b, \mu - \frac{(i + j + s)}{2}\right)[-t] \tag{6.16}$$

Now the autocorrelation,

$$\int_{-\infty}^{\infty} u_i^{(t)}\, u_j^{(t-s)}\, \mathrm{d}t = \int_{-\infty}^{\infty} \mathcal{N}\left(a^2\, e^{-\frac{b(s+j-i)^2}{2}}, 2b, \mu - \frac{(i + j + s)}{2}\right)[-t]\, \mathrm{d}t \tag{6.17}$$

Using Lemma 2

$$= a^2 \sqrt{\frac{\pi}{2b}}\, e^{-\frac{b(s-(i-j))^2}{2}} = \mathcal{N}\left(a^2 \sqrt{\frac{\pi}{2b}}, \frac{b}{2}, (i - j)\right)[s] \tag{6.18}$$

Computing $W_{ij}^{(\epsilon=0)}$ (the weight matrix without any symmetric correction) using autocorrelations

$$W_{ij}^{(\epsilon=0)} = -A_- \int_{-\infty}^{0} e^{s/\tau_-} \mathcal{N}\left(a^2\sqrt{\frac{\pi}{2b}}, \frac{b}{2}, (i-j)\right)[s]\,\mathrm{d}s + A_+ \int_{0}^{\infty} e^{-s/\tau_+} \mathcal{N}\left(a^2\sqrt{\frac{\pi}{2b}}, \frac{b}{2}, (i-j)\right)[s]\,\mathrm{d}s \tag{6.19}$$

$$W_{ij}^{(\epsilon=0)} = -A_- a^2\sqrt{\frac{\pi}{2b}} \int_{-\infty}^{0} e^{s/\tau_-} \mathcal{N}\left(1, \frac{b}{2}, (i-j)\right)[s]\,\mathrm{d}s + A_+ a^2\sqrt{\frac{\pi}{2b}} \int_{0}^{\infty} e^{-s/\tau_+} \mathcal{N}\left(1, \frac{b}{2}, (i-j)\right)[s]\,\mathrm{d}s \tag{6.20}$$

using Lemma 4

$$\begin{aligned} W_{ij}^{(\epsilon=0)} = &-A_- a^2\sqrt{\frac{\pi}{2b}} \int_{-\infty}^{0} \mathcal{N}\left(\exp\left(\frac{i-j}{\tau_-} + \frac{1}{b\tau_-^2}\right), \frac{b}{2}, (i-j) + \frac{1}{b\tau_-}\right)[s]\,\mathrm{d}s \\ &+ A_+ a^2\sqrt{\frac{\pi}{2b}} \int_{0}^{\infty} \mathcal{N}\left(\exp\left(-\frac{i-j}{\tau_+} + \frac{1}{b\tau_+^2}\right), \frac{b}{2}, (i-j) - \frac{1}{b\tau_+}\right)[s]\,\mathrm{d}s \end{aligned} \tag{6.21}$$

Let $\Delta = i - j$,

$$W_{ij}^{(\epsilon=0)} = \frac{\pi a^2}{2b}\left[A_+ e^{\left(\frac{1}{b\tau_+^2} - \frac{\Delta}{\tau_+}\right)} \operatorname{erfc}\left(\frac{1 - b\tau_+\Delta}{\tau_+\sqrt{2b}}\right) - A_- e^{\left(\frac{1}{b\tau_-^2} + \frac{\Delta}{\tau_-}\right)} \operatorname{erfc}\left(\frac{1 + b\tau_-\Delta}{\tau_-\sqrt{2b}}\right)\right] \tag{6.22}$$

When the STDP has the symmetric correction (ϵ), we have:

$$\begin{aligned} W_{ij} = &\frac{\pi a^2}{2b}\left[A_+ e^{\left(\frac{1}{b\tau_+^2} - \frac{\Delta}{\tau_+}\right)} \operatorname{erfc}\left(\frac{1 - b\tau_+\Delta}{\tau_+\sqrt{2b}}\right) - A_- e^{\left(\frac{1}{b\tau_-^2} + \frac{\Delta}{\tau_-}\right)} \operatorname{erfc}\left(\frac{1 + b\tau_-\Delta}{\tau_-\sqrt{2b}}\right)\right] \\ &+ \epsilon\frac{\pi a^2}{2b}\left[A_+ e^{\left(\frac{1}{b\tau_+^2} + \frac{\Delta}{\tau_+}\right)} \operatorname{erfc}\left(\frac{1 + b\tau_+\Delta}{\tau_+\sqrt{2b}}\right) - A_- e^{\left(\frac{1}{b\tau_-^2} - \frac{\Delta}{\tau_-}\right)} \operatorname{erfc}\left(\frac{1 - b\tau_-\Delta}{\tau_-\sqrt{2b}}\right)\right] \end{aligned} \tag{6.23}$$

(i) STDP Eigenvalues

Eigenvalues control the dynamics the RNN has; however, they are difficult to compute directly from the equations for W_{ij}. Instead, we compute it from the learning kernel in Equation **??** directly. Thus, the resultant W_{ij} is a Toeplitz matrix since W_{ij} is a function of $\Delta = i - j$, which are asymptotically diagonalized by the DFT matrix. We have

$$W_{ij} = \int K(s) \int u_i^{(t)}\, u_j^{(t-s)}\,\mathrm{d}t\,\mathrm{d}s + \epsilon \int K(s) \int u_j^{(t)}\, u_i^{(t-s)}\,\mathrm{d}t\,\mathrm{d}s \tag{6.24}$$

$$W_{ij} = a^2\sqrt{\frac{\pi}{2b}}\left(\int K(s)\mathcal{N}\left(1, \frac{b}{2}, \Delta\right)[s]\,\mathrm{d}s + \epsilon\int K(s)\mathcal{N}\left(1, \frac{b}{2}, -\Delta\right)[s]\,\mathrm{d}s\right) \tag{6.25}$$

$$W_{ij} = a^2\sqrt{\frac{\pi}{2b}}\left(\int K(s)\mathcal{N}\left(1, \frac{b}{2}, s - \Delta\right)\mathrm{d}s + \epsilon\int K(s)\mathcal{N}\left(1, \frac{b}{2}, s + \Delta\right)\mathrm{d}s\right) \tag{6.26}$$

Using the convolution theorem, we have

$$W_{ij} = a^2\sqrt{\frac{\pi}{2b}}\mathcal{F}^{-1}\left(\mathcal{F}[K(s)]\,\mathcal{F}\left[\mathcal{N}\left(1, \frac{b}{2}, s\right)\right] + \epsilon\mathcal{F}[K(s)]^\dagger\,\mathcal{F}\left[\mathcal{N}\left(1, \frac{b}{2}, s\right)\right]\right) \tag{6.27}$$

$$= a^2\sqrt{\frac{\pi}{2b}}\sqrt{\frac{2\pi}{b}}\mathcal{F}^{-1}\left[e^{\left(-\frac{\omega^2}{2b}\right)}\left(\mathcal{F}[K(s)] + \epsilon\mathcal{F}[K(s)]^\dagger\right)\right] \tag{6.28}$$

$$= a^2\sqrt{\frac{\pi}{2b}}\sqrt{\frac{2\pi}{b}}\mathcal{F}^{-1}\left[\left(\frac{A_+}{\frac{1}{\tau_+} + i\omega} - \frac{A_-}{\frac{1}{\tau_-} - i\omega}\right)e^{\left(-\frac{\omega^2}{2b}\right)}(1+\epsilon)\right] \tag{6.29}$$

$$= \frac{\pi a^2}{b}\mathcal{F}^{-1}\left[e^{\left(-\frac{\omega^2}{2b}\right)}\left((1+\epsilon)\left(\frac{A_+\tau_+}{1+\omega^2\tau_+^2} - \frac{A_-\tau_-}{1+\omega^2\tau_-^2}\right) - \mathbf{i}(1-\epsilon)\left(\frac{A_+\tau_+^2\omega}{1+\omega^2\tau_+^2} - \frac{A_-\tau_-^2\omega}{1+\omega^2\tau_-^2}\right)\right)\right] \tag{6.30}$$

From this, we obtain the eigenvalue expression

$$\lambda(\omega) = \frac{\pi a^2}{b}e^{\left(-\frac{\omega^2}{2b}\right)}\left((1+\epsilon)\left(\frac{A_+\tau_+}{1+\omega^2\tau_+^2} - \frac{A_-\tau_-}{1+\omega^2\tau_-^2}\right) - \mathbf{i}(1-\epsilon)\left(\frac{A_+\tau_+^2\omega}{1+\omega^2\tau_+^2} - \frac{A_-\tau_-^2\omega}{1+\omega^2\tau_-^2}\right)\right) \tag{6.31}$$

(ii) Useful STDP Lemmas

Lemma 1 (Product of Gaussian). *If* $\mathcal{N}(a,b,\mu) = ae^{-b(x-\mu)^2}$, *then*

$$\mathcal{N}(a_1,b_1,\mu_1)\mathcal{N}(a_2,b_2,\mu_1) = \mathcal{N}\left(a_1a_2\, e^{-\frac{b_1b_2(\mu_1-\mu_2)^2}{(b_1+b_2)}}, (b_1+b_2), \frac{b_1\mu_1+b_2\mu_2}{b_1+b_2}\right) \tag{6.32}$$

Proof. Let the product of the two Gaussian functions be defined as $f(x) = \mathcal{N}(a_1,b_1,\mu_1)\mathcal{N}(a_2,b_2,\mu_2)$. By substituting the definition $\mathcal{N}(a,b,\mu) = ae^{-b(x-\mu)^2}$, we obtain:

$$f(x) = a_1a_2 \exp\left(-\left[b_1(x-\mu_1)^2 + b_2(x-\mu_2)^2\right]\right) \tag{6.33}$$

Let $Q(x)$ denote the quadratic form in the exponent. Expanding the terms, we have:

$$\begin{aligned} Q(x) &= b_1(x^2 - 2x\mu_1 + \mu_1^2) + b_2(x^2 - 2x\mu_2 + \mu_2^2) \\ &= (b_1+b_2)x^2 - 2x(b_1\mu_1 + b_2\mu_2) + (b_1\mu_1^2 + b_2\mu_2^2) \end{aligned} \tag{6.34}$$

To express $Q(x)$ in the standard form $b_c(x-\mu_c)^2 + K$, we define the combined precision b_c and combined mean μ_c as:

$$b_c = b_1 + b_2, \quad \mu_c = \frac{b_1\mu_1 + b_2\mu_2}{b_1+b_2} \tag{6.35}$$

Rewriting the quadratic form by completing the square with respect to x:

$$Q(x) = (b_1+b_2)\left(x - \frac{b_1\mu_1+b_2\mu_2}{b_1+b_2}\right)^2 + K \tag{6.36}$$

where the residual constant K is:

$$K = b_1\mu_1^2 + b_2\mu_2^2 - \frac{(b_1\mu_1+b_2\mu_2)^2}{b_1+b_2} \tag{6.37}$$

Simplifying K by finding a common denominator:

$$\begin{aligned} K &= \frac{(b_1\mu_1^2 + b_2\mu_2^2)(b_1+b_2) - (b_1^2\mu_1^2 + 2b_1b_2\mu_1\mu_2 + b_2^2\mu_2^2)}{b_1+b_2} \\ &= \frac{b_1b_2\mu_1^2 + b_1b_2\mu_2^2 - 2b_1b_2\mu_1\mu_2}{b_1+b_2} \\ &= \frac{b_1b_2(\mu_1-\mu_2)^2}{b_1+b_2} \end{aligned} \tag{6.38}$$

Substituting $Q(x) = b_c(x-\mu_c)^2 + K$ back into the exponential function:

$$\begin{aligned} f(x) &= a_1a_2 \exp(-K) \exp\left(-b_c(x-\mu_c)^2\right) \\ &= \left[a_1a_2 \exp\left(-\frac{b_1b_2(\mu_1-\mu_2)^2}{b_1+b_2}\right)\right] \exp\left(-(b_1+b_2)(x-\mu_c)^2\right) \end{aligned} \tag{6.39}$$

By inspection, this result corresponds to the function $\mathcal{N}(a_c,b_c,\mu_c)$. □

Lemma 2 (Gaussian Integral). *If* $\mathcal{N}(a,b,\mu) = ae^{-b(x-\mu)^2}$, *then*

$$\int_{-\infty}^{\infty} \mathcal{N}(a,b,\mu)\,\mathrm{d}x = a\sqrt{\frac{\pi}{b}} \tag{6.40}$$

Proof. Let the integral be denoted by I. We aim to evaluate:

$$I = \int_{-\infty}^{\infty} ae^{-b(x-\mu)^2}\,\mathrm{d}x \tag{6.41}$$

First, we factor out the constant amplitude a:

$$I = a\int_{-\infty}^{\infty} e^{-b(x-\mu)^2}\,\mathrm{d}x \tag{6.42}$$

We perform a change of variables to simplify the exponent. Let $u = \sqrt{b}(x-\mu)$. Consequently, the differential transforms as:

$$\mathrm{d}u = \sqrt{b}\,\mathrm{d}x \quad \Longrightarrow \quad \mathrm{d}x = \frac{1}{\sqrt{b}}\,\mathrm{d}u \tag{6.43}$$

The limits of integration remain from $-\infty$ to ∞. Substituting these into the integral:

$$I = a\int_{-\infty}^{\infty} e^{-u^2}\left(\frac{1}{\sqrt{b}}\right)\,\mathrm{d}u = \frac{a}{\sqrt{b}}\int_{-\infty}^{\infty} e^{-u^2}\,\mathrm{d}u \tag{6.44}$$

We recall the standard Gaussian integral identity:

$$\int_{-\infty}^{\infty} e^{-u^2}\,\mathrm{d}u = \sqrt{\pi} \tag{6.45}$$

Substituting this standard result back into our expression for I:

$$I = \frac{a}{\sqrt{b}}\left(\sqrt{\pi}\right) = a\sqrt{\frac{\pi}{b}} \tag{6.46}$$

This completes the proof. □

Lemma 3 (Gaussian Integral Bounds). *If* $\mathcal{N}(a,b,\mu) = ae^{-b(x-\mu)^2}$, *then*

$$\int_L^U \mathcal{N}(a,b,\mu)\,\mathrm{d}x = a\frac{1}{2}\sqrt{\frac{\pi}{b}}\left[erf\left(\sqrt{b}(U-\mu)\right) - erf\left(\sqrt{b}(L-\mu)\right)\right] \tag{6.47}$$

Proof. Let the integral be denoted by I. We aim to evaluate the definite integral:

$$I = \int_L^U ae^{-b(x-\mu)^2}\,\mathrm{d}x \tag{6.48}$$

First, we factor out the constant amplitude a:

$$I = a\int_L^U e^{-b(x-\mu)^2}\,\mathrm{d}x \tag{6.49}$$

We perform a change of variables to standardize the exponent. Let $t = \sqrt{b}(x-\mu)$. The differential transforms as $\mathrm{d}x = \frac{1}{\sqrt{b}}\,\mathrm{d}t$. We must also transform the limits of integration:

- When $x = L$, $t_L = \sqrt{b}(L-\mu)$.
- When $x = U$, $t_U = \sqrt{b}(U-\mu)$.

Substituting these into the integral:

$$I = \frac{a}{\sqrt{b}}\int_{t_L}^{t_U} e^{-t^2}\,\mathrm{d}t \tag{6.50}$$

We recall the definition of the error function, $\mathrm{erf}(z)$, which is defined as:

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,\mathrm{d}t \tag{6.51}$$

Using the property of definite integrals that $\int_A^B f(t)\,\mathrm{d}t = \int_0^B f(t)\,\mathrm{d}t - \int_0^A f(t)\,\mathrm{d}t$, we can rewrite our integral in terms of the error function:

$$\int_{t_L}^{t_U} e^{-t^2}\,\mathrm{d}t = \left(\int_0^{t_U} e^{-t^2}\,\mathrm{d}t - \int_0^{t_L} e^{-t^2}\,\mathrm{d}t\right) \tag{6.52}$$

From the definition of $\mathrm{erf}(z)$, we have $\int_0^z e^{-t^2}\,\mathrm{d}t = \frac{\sqrt{\pi}}{2}\mathrm{erf}(z)$. Therefore:

$$\int_{t_L}^{t_U} e^{-t^2}\,\mathrm{d}t = \frac{\sqrt{\pi}}{2}\mathrm{erf}(t_U) - \frac{\sqrt{\pi}}{2}\mathrm{erf}(t_L) = \frac{\sqrt{\pi}}{2}\left[\mathrm{erf}(t_U) - \mathrm{erf}(t_L)\right] \tag{6.53}$$

Substituting this result back into the expression for I:

$$I = \frac{a}{\sqrt{b}}\left(\frac{\sqrt{\pi}}{2}\left[\mathrm{erf}(t_U) - \mathrm{erf}(t_L)\right]\right) \tag{6.54}$$

Rearranging terms and substituting the definitions of t_U and t_L:

$$I = a\frac{1}{2}\sqrt{\frac{\pi}{b}}\left[\mathrm{erf}\left(\sqrt{b}(U-\mu)\right) - \mathrm{erf}\left(\sqrt{b}(L-\mu)\right)\right] \tag{6.55}$$

□

Lemma 4 (Gaussian Exp Product). *If* $\mathcal{N}(a,b,\mu) = ae^{-b(x-\mu)^2}$, *then*

$$e^{cx}\,\mathcal{N}(a,b,\mu) = \mathcal{N}\left(a\exp\left(c\mu + \frac{c^2}{4b}\right), b, \mu + \frac{c}{2b}\right) \tag{6.56}$$

Proof. Let the product function be $f(x) = e^{cx}\,\mathcal{N}(a, b, \mu)$. Substituting the definition of the Gaussian function:

$$f(x) = e^{cx} \cdot ae^{-b(x-\mu)^2} = a\exp\left(cx - b(x-\mu)^2\right) \tag{6.57}$$

Let $E(x)$ denote the exponent. We expand the quadratic term to group coefficients of x:

$$\begin{aligned} E(x) &= cx - b(x^2 - 2x\mu + \mu^2) \\ &= -bx^2 + 2bx\mu + cx - b\mu^2 \\ &= -bx^2 + (2b\mu + c)x - b\mu^2 \end{aligned} \tag{6.58}$$

To restore the Gaussian form $-b(x-\mu_{\text{new}})^2 + K$, we factor out $-b$ from the terms involving x:

$$E(x) = -b\left(x^2 - \frac{2b\mu + c}{b}x\right) - b\mu^2 \tag{6.59}$$

We define the new mean μ_{new} as half the coefficient of the linear term inside the parenthesis:

$$\mu_{\text{new}} = \frac{1}{2}\left(\frac{2b\mu + c}{b}\right) = \mu + \frac{c}{2b} \tag{6.60}$$

Completing the square inside the parenthesis by adding and subtracting μ_{new}^2:

$$\begin{aligned} E(x) &= -b\left(x^2 - 2x\mu_{\text{new}} + \mu_{\text{new}}^2 - \mu_{\text{new}}^2\right) - b\mu^2 \\ &= -b(x-\mu_{\text{new}})^2 + b\mu_{\text{new}}^2 - b\mu^2 \end{aligned} \tag{6.61}$$

We simplify the constant terms (the residual) $K = b\mu_{\text{new}}^2 - b\mu^2$:

$$\begin{aligned} K &= b\left(\mu + \frac{c}{2b}\right)^2 - b\mu^2 \\ &= b\left(\mu^2 + \frac{2\mu c}{2b} + \frac{c^2}{4b^2}\right) - b\mu^2 \\ &= b\mu^2 + c\mu + \frac{c^2}{4b} - b\mu^2 \\ &= c\mu + \frac{c^2}{4b} \end{aligned} \tag{6.62}$$

Substituting $E(x) = -b(x-\mu_{\text{new}})^2 + K$ back into the original expression for $f(x)$:

$$f(x) = a\exp(K)\exp\left(-b(x-\mu_{\text{new}})^2\right) \tag{6.63}$$

Substituting the values of K and μ_{new}:

$$f(x) = \left[a\exp\left(c\mu + \frac{c^2}{4b}\right)\right]\exp\left(-b\left(x - \left(\mu + \frac{c}{2b}\right)\right)^2\right) \tag{6.64}$$

This matches the definition of $\mathcal{N}(a_{\text{new}}, b, \mu_{\text{new}})$. □