Memory flows: geometry and dynamics of sequential retrieval in input-driven Hopfield networks

Simone Betteti
Robust and Intelligent Autonomous Systems (RIAS) lab
The Italian Institute of Artificial Intelligence (AI4I)
simone.betteti@ai4i.it

Matteo Martin
Department of Information Engineering (DEI)
University of Padova
matteo.martin.2@phd.unipd.it

Morten Gram Pedersen
Department of Information Engineering (DEI)
Padua Neuroscience Center (PNC)
University of Padova
mortengram.pedersen@unipd.it

Sandro Zampieri
Department of Information Engineering (DEI)
University of Padova
zampi@dei.unipd.it

Giacomo Baggio
Department of Information Engineering (DEI)
University of Padova
baggio@dei.unipd.it

Abstract

Associative memory models classically describe retrieval as convergence to stored patterns, but the theory of sequential retrieval remains comparatively limited. We develop a two-timescale input-driven Hopfield network in which fast associative retrieval is modulated by a slow feedback variable evolving over a prescribed memory-transition graph. This yields autonomous, graph-constrained transitions while preserving analytical tractability. Using slow-fast and geometric singular perturbation theory, we characterize memory fixation via memory manifolds and identify the geometric regions that govern stability loss and escape from retrieved memories. We further derive reduced slow dynamics and explicit criteria for transition onset, self-sustained retrieval, and collapse, expressed through escape times, gain thresholds, and fixed points. Together, these results provide a tractable framework for analyzing feedback-driven sequential retrieval in Hopfield networks.

1 Introduction

Associative memory has been a foundational concept in theoretical neuroscience and machine learning for over four decades. In its classical formulation, associative memory retrieval is modeled as convergence toward stable equilibria of an energy landscape [1–3], whereby a corrupted cue relaxes to a stored pattern through recurrent dynamics, providing a canonical example of distributed computation in neural networks. A large body of work has focused on characterizing the memory capacity of such models [4–6], revealing a fundamental limitation of the original Hopfield network, whose capacity scales sub-linearly with the number of neurons. Modern Hopfield models [7–9] overcome this limitation by achieving exponential capacity within an energy-based framework, and have been linked to contemporary architectures such as Transformers [10, 11] and diffusion models [12, 13]. However, they largely remain static retrieval systems: dynamics stop at convergence, so sequential retrieval requires repeated reinitialization. This contrasts with cognitive and algorithmic

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

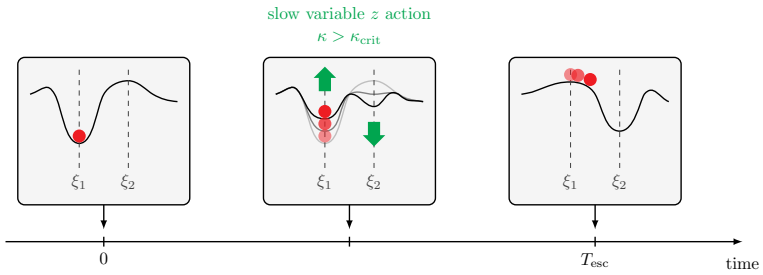

Figure 1: *Schematic illustration of the proposed two-timescale mechanism for sequential memory transitions.* At the initial time, the fast state variable is at a stable equilibrium corresponding to memory ξ_1. The slow variable z progressively destabilizes this equilibrium while simultaneously stabilizing the equilibrium associated with memory ξ_2. At the escape time T_{esc}, the equilibrium corresponding to ξ_1 loses stability and the system transitions toward ξ_2. This transition occurs only when the gain parameter κ of the slow dynamics exceeds a critical threshold κ_{crit}. We analytically characterize both κ_{crit} and the resulting escape time T_{esc}.

tasks, including those addressed by Transformers, where reasoning depends on structured transitions across sequences of memories.

Pioneering works by [14, 15] already identified the limitation of a theory accounting only for static retrieval. They proposed a continuous-time mechanism for sequential memory transitions via delayed asymmetric interactions, which introduce a destabilizing dynamic component driving transitions between attractors. This idea has since motivated extensive efforts to model memory transitions [16, 17], establishing mean-field results on the gain sustaining sequentiality. More recently, [18] introduced a continuous-time, slow-fast energy-based model with delayed asymmetric interactions, while [19] investigated long-sequence retrieval and capacity in a discrete-time setting. The softmax-based model of [20] further provides capacity results and escape-time estimates based primarily on energy arguments. Related approaches use slowly evolving parameters to control attractor switching [21, 22]. Our aim is complementary: rather than improving retrieval performance, we focus on developing an analytically tractable model that provides a mechanistic understanding of sequential memory transitions.

In this paper, we propose and analyze a two-timescale extension of the input-driven plasticity (IDP) Hopfield model of [23] that enables autonomous sequential retrieval. In the original IDP Hopfield model, external inputs modulate the saliency of stored memories and thereby reshape the retrieval landscape. Here, we replace this externally prescribed modulation with a slow feedback variable whose dynamics are driven by the currently retrieved memory and structured by a transition graph. The slow feedback can be viewed as a phenomenological representation of activity-dependent biological processes, such as short-term synaptic plasticity, neuronal adaptation, or neuromodulatory gain control, that modulate memory stability [24]. The resulting sequential IDP Hopfield network produces graph-constrained transitions between memories while remaining analytically tractable. Our contributions are threefold. First, using slow-fast and geometric singular perturbation theory, we characterize the geometry of memory fixation through memory manifolds and the attracting and saddle-type regions that govern entry and escape. Second, we derive reduced saliency dynamics and explicit escape conditions, including transition times, gain thresholds, and collapse regimes. Third, we show how these conditions support self-sustained sequential retrieval over structured memory graphs. Fig. 1 provides a schematic overview of the sequential memory transition mechanism introduced in this work. Together, these results establish a tractable geometric and dynamical framework for sequential retrieval in input-driven Hopfield networks, providing explicit conditions under which slow internal feedback induces, sustains, or fails to sustain transitions between memories.

2 The IDP Hopfield model for sequential retrieval

The IDP Hopfield model introduced in [23] extends the classical Hopfield model by incorporating an explicit modulation of synaptic interactions through external inputs. This formulation enables a controlled and interpretable mechanism for memory retrieval, where the input does not merely select an initial condition but directly shapes the energy landscape governing the dynamics. We next

briefly review the IDP Hopfield model, deferring technical details to Appendix A, and then introduce a sequential extension that enables autonomous, structured transitions between stored memories.

The IDP Hopfield model. The IDP Hopfield dynamics, illustrated schematically in Fig. 2(A), are

$$\dot{x} = -x + W\big(\Phi(u)\big)\Psi(x), \qquad x(0) \in \mathbb{R}^N, \tag{1}$$

where $x \in \mathbb{R}^N$ is the network state and $u \in \mathbb{R}^d$ is a slowly varying external input. The map $\Phi : \mathbb{R}^d \to \mathbb{R}^P_{\geq 0}$ is a *saliency function* that acts as a positive filter on the input u, while $\Psi : \mathbb{R}^N \to \mathbb{R}^N$ is the neural *activation function*. The synaptic matrix depends on the filtered input through

$$W(\Phi(u)) = M\mathrm{diag}(\Phi(u))M^\top, \qquad M := \tfrac{1}{\sqrt{N}}[\xi^1, \dots, \xi^P],$$

where $\{\xi^\mu\}_{\mu=1}^P \subseteq \{\pm 1\}^N$ is a set of mutually orthogonal stored patterns. We stress that the orthogonality constraint is necessary to derive rigorous results on the geometry and dynamics of the discussed model, but numerical experiments support their validity also in the canonical case of random memories.

Throughout the paper, we assume that $\Phi(\cdot)$ acts componentwise and homogeneously, i.e., $\Phi_i(u) = \phi(u_i)$ for some scalar function $\phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$, which is even and monotone with respect to $|u_i|$. Similarly, $\Psi(\cdot)$ acts componentwise as $\Psi_i(x) = \psi(x_i)$, where $\psi : \mathbb{R} \to \mathbb{R}$ is odd, continuously-differentiable and monotone non-decreasing (e.g., $\tanh$). Under these assumptions, the map $\Phi(\cdot)$ assigns a non-negative saliency coefficient to each stored pattern, thereby modulating its contribution to the synaptic matrix $W(\Phi(u))$, while $\Psi(\cdot)$ retains the standard symmetry and monotonicity properties of Hopfield-type activation functions. The classical Hopfield model is recovered as a special case when all input-dependent coefficients $\phi(u_i)$, $i = 1, \dots, P$, are all equal to one, in which case the synaptic matrix reduces to a fixed matrix independent of the input.

Under the above-mentioned standard assumptions on Ψ, the IDP Hopfield model admits the energy function [3, 8]

$$\mathrm{E}\big(x, \Phi(u)\big) = -\tfrac{1}{2}\Psi(x)^\top W\big(\Phi(u)\big)\Psi(x) + x^\top \Psi(x) - \sum_{i=1}^N \int_0^{x_i} \psi(s)\, ds. \tag{2}$$

This energy provides a geometric interpretation of retrieval: the input-dependent saliencies $\Phi(u)$ reshape the energy landscape, thereby controlling the existence and stability of memory equilibria.

A key feature of the IDP formulation is that the input-dependent coefficients $\{\phi(u_\mu)\}_{\mu=1}^P$ directly control the existence and stability of memory states. In general, the system admits multiple equilibria, some of which are associated with the stored patterns.[1] In particular, to each memory vector ξ^ν one can associate a *memory equilibrium* of the form $x = \gamma_\nu \xi^\nu$, which exists if and only if $\phi(u_\nu) > 1$. In that case, the scaling factor γ_ν is uniquely determined by $\gamma_\nu = f^{-1}(\phi(u_\nu))$, where $f(\gamma) := \gamma/\psi(\gamma)$. The stability of these equilibria is also governed by the coefficients $\phi(u_\nu)$: if $\phi(u_\nu) > 1$, the corresponding equilibrium is stable only if $\phi(u_\nu)$ is sufficiently large relative to the maximum coefficient $\phi_{\max} := \max_\mu \phi(u_\mu)$ (see Appendix A for details). Overall, the coefficients $\phi(u_\mu)$ provide a direct and tunable mechanism to determine which memories exist and are stable, thereby enabling selective activation and suppression within the same network.

The Two-Timescale Sequential IDP Hopfield Model. While the IDP formulation enables input-modulated retrieval and sharp control over memory stability, it remains fundamentally static in nature: for a fixed input, the dynamics converge to a single equilibrium and do not exhibit autonomous transitions between memories. In particular, the model lacks an internal mechanism to encode temporal or semantic relationships among stored patterns.

To overcome this limitation, we introduce a minimal two-timescale extension of (1) that couples the fast retrieval dynamics to a slower internal state:

$$\begin{cases} \tau_x \dot{x} = -x + M\mathrm{diag}\big(\Phi(z)\big)M^\top \Psi(x), & x(0) \in \mathbb{R}^N, \\ \tau_z \dot{z} = -z + \frac{\kappa}{\sqrt{N}} A M^\top \Psi(x), & z(0) \in \mathbb{R}^P, \end{cases} \tag{3}$$

where $\kappa > 0$ is a gain parameter and $A \in \{0,1\}^{P \times P}$ encodes structured associations among memories. Specifically, A is the adjacency matrix of a directed graph $\mathcal{G}_A = (\mathcal{V}, \mathcal{E})$, whose nodes

[1] Other equilibria (so-called *spurious* states) may exist, but they do not play a relevant role in the setting considered here.

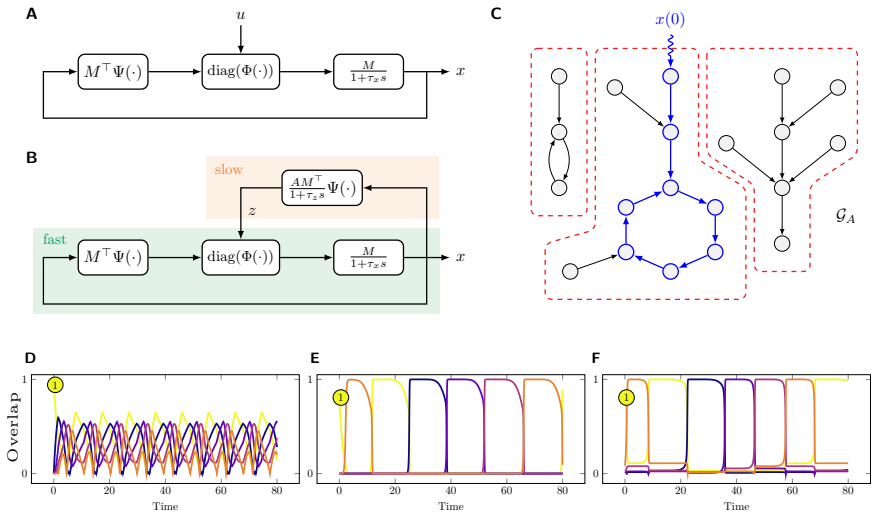

Figure 2: (A) Block diagram of the IDP Hopfield model; setting $\tau_x = 1$ recovers (1). (B) Block diagram of the two-timescale sequential IDP Hopfield model (3). (C) Example of a deterministic transition graph $\mathcal{G}_A$, where nodes represent stored memories and edges encode admissible transitions. The graph decomposes into multiple connected components (highlighted by dashed red contours), each defining a distinct memory sequence; one, selected by $x(0)$, is highlighted in blue. (D) One-timescale model with orthogonal memories ($\kappa = 5$): sequential retrieval proceeds through mixed memory states, with partial overlaps and unpredictable escape times across memories. (E-F) Two-timescale sequential IDP Hopfield model ($\tau_x = 1, \tau_z = 5, \kappa = 5$). (E) Orthogonal memories: transitions occur between individual memories without mixing, with uniform escape times and overlaps reaching their maximal value, indicating exact memory alignment. (F) Random memories: transitions occur without mixing, reaching maximum overlap, but escape times are no longer uniform. Numerical experiments are shown for visualization purposes with $P = 15$ memories and a limit cycle over 5 of them, with $\psi = \tanh$, and whose overlaps are highlighted in different colors in the plots.

$\mathcal{V} = \{1, \dots, P\}$ represent stored memories and whose edges represent admissible transitions, with $(i, j) \in \mathcal{E}$ if and only if $A_{ji} = 1$. We focus on deterministic graphs, where each node has out-degree at most one, and assume $\tau_x \ll \tau_z$, so that x evolves rapidly while z provides slow modulation.

Fig. 2(B) illustrates the block diagram of the sequential IDP Hopfield model (3). In contrast to the original IDP formulation (1), where the input directly modulates the synaptic weights, here the modulation is mediated by the slow, feedback-driven variable z, which endows the system with an internal memory of past retrieval states. Fig. 2(C) shows an example of deterministic graph $\mathcal{G}_A$. Such graphs may consist of multiple connected components, each corresponding to a distinct memory sequence that can be selectively retrieved through appropriate choice of the initial condition $x(0)$.

In the regime $\tau_x \ll \tau_z$, a clear separation of timescales emerges in (3): the fast variable x rapidly relaxes toward a memory determined by the current $\Phi(z)$, while the slow variable z evolves in response to the retrieved pattern and progressively reshapes the energy landscape. This interaction gives rise to autonomous sequential retrieval, whereby the system transitions between stored memories according to the graph $\mathcal{G}_A$. As illustrated in Fig. 2(E-F) on a simple cyclic graph with $P = 5$ memories, the proposed two-timescale sequential IDP Hopfield model exhibits robust and predictable sequential retrieval when κ is sufficiently large. For orthogonal memories (Fig. 2(E)), transitions occur directly between individual memories without intermediate mixing: overlaps reach their maximal value, indicating exact memory alignment, and escape times are uniform across memories. For random memories (Fig. 2(F)), the dynamics similarly transition without mixing and attain maximal overlap, although escape times are no longer uniform. To highlight the role of timescale separation, we compare this behavior with the classical sequential Hopfield dynamics obtained in a single-timescale setting, $\dot{x} = -x + \kappa M A M^\top \Psi(x)$, where transitions are encoded directly in the synaptic matrix [25]. As shown in Fig. 2(D), this model exhibits qualitatively different dynamics: trajectories evolve through mixed states with partial overlaps, display irregular escape times, and are highly sensitive to

the gain κ. Depending on its value, the system may collapse to the origin, converge to a fixed point, or sustain a limit cycle, consistent with previous observations.

Finally, we point out that (3) admits a natural modern Hopfield reformulation, detailed in Appendix B.

3 Memory manifolds: the highway of sequential retrieval

Sequential memory retrieval is the continuous evolution of neural activity in state space, from one stable memory state to another one. The direction of motion is shaped by the semantic relationships between memories, which at any given time constrain neural activity to a lower-dimensional memory subspace. As semantic relationships change, the binding is broken, allowing neural trajectories to leave one memory subspace and evolve toward another.

From a geometric perspective, as sketched in Fig. 3(A), sequential retrieval therefore requires two complementary structures: (i) *memory subspaces* (red and shaded surfaces) that transiently trap neural activity (blue curves), and (ii) *entry and exit pathways* that enable transitions (purple branches) between them. In this section, we formalize this intuition using slow-fast and *geometric singular perturbation theory* (GSPT) [26, 27]. Leveraging the analytical tractability of the IDP Hopfield model, we provide a geometric characterization of how sequential retrieval arises in high-dimensional associative memory systems.

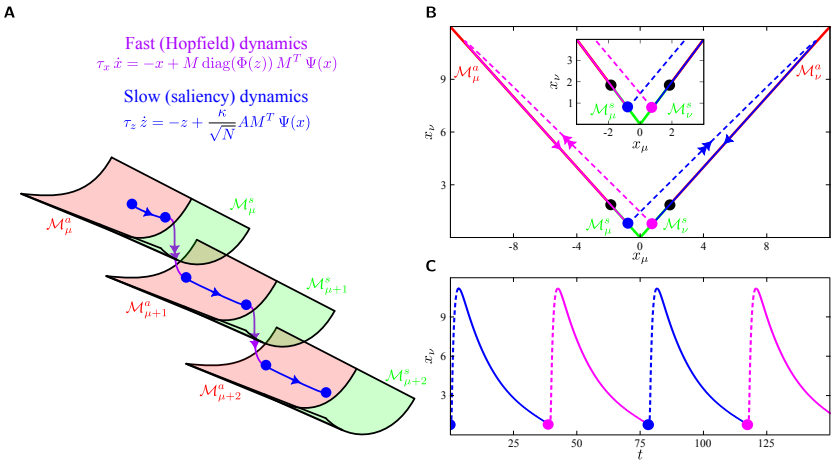

Figure 3: (A) Sketch of the transitions between different memory manifolds in the slow-fast Hopfield model. The blue curves corresponds to the slow movement of the system close to a stable memory manifold, whereas the purple curves correspond to the fast switching between different stable memory manifolds. Each memory manifold $\mathcal{N}_\mu$ is decomposed into its attracting (red) and saddle-type (green) memory manifolds $\mathcal{M}^a_\mu$, $\mathcal{M}^s_\mu$. (B) Simulated trajectory projected onto the fast subspace in a slow-fast Hopfield model storing two memories. Continuous curves indicate the flow along memory branches $\mathcal{N}_\mu$ (magenta) and $\mathcal{N}_\nu$ (blue), whereas dashed curves indicate rapid transitions from one memory branch to another. Switching occurs after the trajectory has moved from the attracting part of the memory branches $\mathcal{M}^{a,\varepsilon}_\nu$ and $\mathcal{M}^{a,\varepsilon}_\mu$, respectively, to the saddle-type section $\mathcal{M}^{s,\varepsilon}_\nu$ and $\mathcal{M}^{s,\varepsilon}_\mu$. The black dots indicate the relevant point of separation between $\mathcal{M}^a_\nu$ and $\mathcal{M}^s_\nu$, and $\mathcal{M}^a_\mu$ and $\mathcal{M}^s_\mu$, respectively (red/green). (C) Temporal series of the variable x_ν. Colors and line types as in panel B.

Geometric singular perturbation theory [26] provides a theoretical framework to analyze dynamical systems with well-separated time scales. The central idea is that, when fast and slow variables interact weakly, the full dynamics can be approximated as a sequence of slow movements, solutions of a simpler system, interspersed by fast jumps. We introduce the time-scale parameter $\varepsilon = \tau_x/\tau_z$ and consider the slow-fast system

$$\begin{cases} \varepsilon \dot{x} = -x + M\,\mathrm{diag}(\Phi(z))\,M^\top \Psi(x), & x(0) \in \mathbb{R}^N, \\ \dot{z} = -z + \frac{\kappa}{\sqrt{N}} A M^\top \Psi(x), & z(0) \in \mathbb{R}^P. \end{cases} \tag{4}$$

We refer to the x-ODE with the term *fast subsystem* and denote with $F(x, z)$ its right-hand side. Taking the limit $\varepsilon \to 0^+$ constrains the fast dynamics onto $\mathbb{0}_N = F(x, z)$, yielding the *reduced*

system. The algebraic constraint effectively tracks the minimum over the energy as the slow neural feedback evolves, and admits multiple solutions, reflecting that Hopfield networks are designed to store multiple stable patterns. In slow-fast theory, the set of the fast-subsystem equilibria is called *critical set*, and defined as

$$\mathcal{S}_0 = \left\{(x,z) \in \mathbb{R}^N \times \mathbb{R}^P : \mathbb{0}_N = F(x,z)\right\}. \tag{5}$$

Memory branches, memory manifolds, and stability. Given the fast dynamics with P stored memories, a *memory branch* $\mathcal{N}_\nu \subseteq \mathcal{S}_0$ describes a P-dimensional subspace in which neural activity can be temporarily confined. In the IDP Hopfield model, these memory branches take the explicit form

$$\mathcal{N}_\nu = \left\{(x,z) \in \mathbb{R}^N \times \mathbb{R}^P : \exists \gamma_\nu \neq 0,\ x = \gamma_\nu \xi^\nu,\ \phi(z_\nu) = f(\gamma_\nu)\right\}, \tag{6}$$

for $\nu = 1, \dots, P$ and $f(\gamma) = \gamma/\psi(\gamma)$. The *normally hyperbolic* subsets $\mathcal{M}^a_\nu$ and $\mathcal{M}^s_\nu$ of $\mathcal{N}_\nu$ are known as the attracting and saddle-type *memory manifolds*. They can be identified by studying the stability of the fast subsystem equilibria in $\mathcal{N}_\nu$, which are related to the saliency weights. Another relevant *branch* of $\mathcal{S}_0$ is the *origin manifold*,

$$\mathcal{M}_0 = \left\{(x,z) \in \mathbb{R}^N \times \mathbb{R}^P : x = 0\right\}. \tag{7}$$

GSPT ensures that, for $0 < \varepsilon \ll 1$, each $\mathcal{N}_\nu$ ($\mathcal{M}^a_\nu$, $\mathcal{M}^s_\nu$, and $\mathcal{M}_0$) perturbs to a nearby invariant manifold $\mathcal{N}^\varepsilon_\nu$ ($\mathcal{M}^{a,\varepsilon}_\nu$, $\mathcal{M}^{s,\varepsilon}_\nu$, and $\mathcal{M}^\varepsilon_0$) of the full system (4). In particular, the manifolds $\mathcal{M}^{a,\varepsilon}_\nu$ and $\mathcal{M}^{s,\varepsilon}_\nu$ provide a precise geometric description of *memory fixation* and *transitions*.

Fixation, loss of stability, and transitions. Sequential retrieval requires understanding how trajectories *enter* and *leave* $\mathcal{N}^\varepsilon_\nu$, which consists in studying the stability of its constitutive memory manifolds. As presented in Fig. 3(B-C), during *retrieval* of the memory ξ^ν, the neural activity converges to $\mathcal{M}^{a,\varepsilon}_\nu$ rapidly, fixation occurs, and escape is impossible. As the saliency $\phi(z_\nu)$ decreases, an escape direction appears when $\phi(z_\nu)$ crosses the threshold $\phi_{\max}$ from above. Geometrically, this corresponds to crossing (an ε-perturbation of) a non-hyperbolic point on $\mathcal{N}_\nu$ (black circle in Fig. 3(B)), corresponding to the trajectory moving from the attracting manifold $\mathcal{M}^{a,\varepsilon}_\nu$ to the saddle manifold $\mathcal{M}^{s,\varepsilon}_\nu$. Beyond this point, a typical trajectory is ejected via an unstable fiber of $\mathcal{M}^{s,\varepsilon}_\nu$ towards a new attracting manifold $\mathcal{M}^{a,\varepsilon}_\mu$ associated to the memory ξ^μ. However, the jump towards a different memory can occur only if the associated $\phi(z_\mu) > 1$. Otherwise, if for all μ it holds $\phi(z_\mu) < 1$ then the only attracting manifold is $\mathcal{M}^\varepsilon_0$, which drives the full system towards the origin.

4 Slow saliency dynamics: the driving engine

In Sec. 3, we characterized the geometric structure underlying memory retrieval, including the manifolds that support retrieval and the instabilities and fibers that enable transitions between them. We now turn to how the reduced slow dynamics in (4) govern sequential transitions as a function of the gain parameter κ, the initial condition $z(0)$, and the structure of the transition matrix A. We present the main results below, deferring technical derivations and proofs to Appendix D.

Reduced dynamics and retrieval regimes. Consider the sequential IDP Hopfield model (3) in the limit $\tau_x/\tau_z \to 0^+$. The fast variable x rapidly relaxes to a stable equilibrium $x^*(z)$ determined by the current saliency z. Substitution into the slow equation of (3) gives

$$\dot{z} = -z + \frac{\kappa}{\sqrt{N}}\, A M^\top \Psi\big(x^*(z)\big). \tag{8}$$

This reduced equation describes the slow flow along the branches $\{\mathcal{N}_\mu\}_{\mu=1}^P$ of the fast Hopfield subsystem: for each fixed z, the fast dynamics select an equilibrium $x^*_\nu(z)$ and sticks to the associated attracting manifold $\mathcal{M}^a_\nu$ until it loses stability. Only then the trajectory may jump to another memory branch, as pictorially depicted in the introductory Fig. 1.

By the geometric picture of Sec. 3, the saliency state z selects which memory branches contain attracting manifolds and can therefore support fixation. This gives three regimes:

(i) **No active memories.** If $\phi(z_\mu) < 1$ for all μ, no memory equilibrium exists and trajectories converge to the origin by evolving in $\mathcal{M}_0$.

(ii) **Unique active memory.** If there is a unique ν such that $\phi(z_\nu) > 1$, there is only active memory equilibrium, and trajectories rapidly converge to $\mathcal{M}^a_\nu$. This equilibrium is stable and given by

$$x^*(z) = f^{-1}\big(\phi(z_\nu)\big)\,\xi^\nu, \quad \text{where } f(\gamma) = \gamma/\psi(\gamma).$$

(iii) **Competing memories.** If several components $\phi(z_\nu)$ exceed the threshold, multiple memory equilibria may coexist, allowing transitions between the corresponding memory states.

Two-memory reduction. We next restrict to the case in which at most two memories compete. Specifically, we assume that, along the relevant portion of the trajectory, only two components of $z(t)$ are significantly different from zero. Without loss of generality, let the initial time be $t = 0$ and assume $z(0) = Z\boldsymbol{e}^{(\nu)}$, where $\boldsymbol{e}^{(\ell)}$ denotes the ℓ-th vector of the canonical basis of $\mathbb{R}^P$ and $Z > 0$ is the initial value of z_ν.[2]

If $\phi(Z) < 1$, no memory is active: the fast state converges to $x(t) = 0$, and the slow dynamics reduce to $\dot{z} = -z$, driving all saliencies exponentially to zero in $\mathcal{M}_0$. We therefore focus on $\phi(Z) > 1$, where ξ^ν is initially active and the fast state is trapped on its associated memory equilibrium,

$$x^\star_\nu(t) = f^{-1}\big(\phi(z_\nu(t))\big)\,\xi^\nu.$$

In this case, the system initially evolves on $\mathcal{M}^a_\nu$, and the reduced saliency dynamics (8) satisfy

$$\dot{z} = -z + \kappa\,\psi\left(f^{-1}\big(\phi(z_\nu)\big)\right)\boldsymbol{e}^{(\mu)}, \tag{9}$$

where μ is the unique successor of ν in the transition graph $\mathcal{G}_A$. Hence $\dot{z}_\sigma = -z_\sigma$ for all $\sigma \neq \mu$. Since only $z_\nu(0)$ is nonzero, all components except z_ν and z_μ remain identically zero, with $z_\nu(t) = Ze^{-t}$. The successor saliency component satisfies

$$\dot{z}_\mu = -z_\mu + \kappa\, g\big(Ze^{-t}\big), \qquad g := \psi \circ f^{-1} \circ \phi.$$

Note that, on the interval where the source memory exists, $g(Ze^{-t}) > 0$, so z_μ is driven by a positive input. Thus, the transition from ξ^ν to ξ^μ reduces to a two-dimensional saliency dynamics: a decaying source component and a positively driven successor component.

Transition criterion and induced map. We now determine whether the two-dimensional saliency dynamics produces a transition to the successor memory or decays to inactivity. Let $T_{\max}$ be the time at which the source memory ceases to exist, namely $\phi\big(z_\nu(T_{\max})\big) = 1$. Since $z_\nu(t) = Ze^{-t}$, the source memory exists for $t < T_{\max}$ and disappears for $t > T_{\max}$. The outcome is determined by whether the successor saliency $z_\mu(t)$ becomes large enough before this loss of existence occurs. Two possible cases arise (see also Fig. 6 of the Appendix for a graphical illustration).

Case 1: no transition. Assume that

$$z_\mu(t) < z_\nu(t) \qquad \forall\, t \in [0, T_{\max}).$$

Then the successor component never catches up with the source component before the source memory disappears. Hence, after $T_{\max}$, both saliencies fall below the existence threshold and the system enters the inactive regime $\mathcal{M}_0$, where the origin is the only stable equilibrium.

Case 2: transition. Assume instead that there exists a first time $\bar{T} \in [0, T_{\max})$ such that

$$z_\mu(\bar{T}) = z_\nu(\bar{T}).$$

In this case, the following result characterizes the loss of stability of the source branch from $\mathcal{M}^a_\nu$ to $\mathcal{M}^s_\nu$, and the induced update of the successor saliency. The proof is given in Appendix D.

Theorem 1 (Escape condition and saliency update)**.** *Under the above condition, there exists an escape time $T_{\mathrm{esc}} \in [\bar{T}, T_{\max})$ at which the equilibrium of the fast Hopfield subsystem associated with the source memory ξ^ν loses stability. Moreover, T_{esc} is the smallest solution in $[0, T_{\max})$ of*

$$\kappa \int_0^{T_{\mathrm{esc}}} e^{-(T_{\mathrm{esc}}-\tau)} g\big(Ze^{-\tau}\big)\, d\tau = h\big(Ze^{-T_{\mathrm{esc}}}\big), \tag{10}$$

where $h(z) := \phi^{-1}\Big(\frac{1}{\psi'(f^{-1}(\phi(z)))}\Big)$.[3] *At escape, the successor saliency is*

$$Z^+ := z_\mu(T_{\mathrm{esc}}) = h\big(Ze^{-T_{\mathrm{esc}}}\big). \tag{11}$$

[2] We take $Z > 0$ for simplicity. The analysis can be easily extended to $Z \in \mathbb{R}$. We note also that if Z were negative, the transition update below yields $Z^+ > 0$, and all subsequent iterations have non-negative amplitudes.

[3] Here ϕ^{-1} is taken on $\mathbb{R}_{\geq 0}$, where ϕ is invertible by assumption.

At time T_{esc}, the trajectory leaves the branch associated with the source memory ξ^ν and jumps to the manifold $\mathcal{M}^a_\mu$ associated with the successor memory ξ^μ. Theorem 1 therefore defines a one-dimensional saliency update $Z \mapsto Z^+$.

For later use, it is convenient to rewrite this update in terms of $X := e^{T_{\text{esc}}}$. Since

$$\int_0^{T_{\text{esc}}} e^{-(T_{\text{esc}}-\tau)} g\big(Ze^{-\tau}\big)\, d\tau = \frac{1}{X}\int_1^X g\left(\frac{Z}{s}\right) ds,$$

the transition map is equivalently described by[4]

$$\begin{cases} R(Z,X) = 0, \\ Z^+ = h(Z/X), \end{cases} \qquad X \in [1, Z/\phi^{-1}(1)), \tag{12}$$

where

$$R(Z,X) := \frac{\kappa}{X}\int_1^X g\left(\frac{Z}{s}\right) ds - h\left(\frac{Z}{X}\right). \tag{13}$$

Thus, each saliency update is computed by first selecting the smallest $X \in [1, Z/\phi^{-1}(1))$ satisfying $R(Z,X) = 0$, when such a solution exists, and then setting $Z^+ = h(Z/X)$. If no such X exists, the transition fails and the trajectory collapses to the inactive state.

5 Analytical illustration and numerical validation

We illustrate the results of Sec. 4 with an analytically tractable example and further assess their validity and robustness through numerical experiments. See also Appendix I.

We consider the saliency update (12)-(13) with the following nonlinearities

$$\psi(x) = \frac{x}{1+|x|}, \qquad \phi(z) = |z|. \tag{14}$$

In particular, we will refer to such ψ as SmoothF. On the positive branch $z \geq 1$, we have $f(\gamma) = \gamma/\psi(\gamma) = \gamma+1$, hence $f^{-1}(z) = z-1$. Therefore $g(z) = \psi(f^{-1}(\phi(z))) = 1-1/z$ and $h(z) = z^2$ for $z \geq 1$. Substituting into (13), for $Z > 1$ and $X \in [1, Z)$,

$$\int_1^X g\left(\frac{Z}{s}\right) ds = (X-1)\left(1 - \frac{X+1}{2Z}\right).$$

Hence the equation $R(Z,X) = 0$ is equivalent to

$$G(Z,X) = \frac{1}{\kappa}, \qquad G(Z,X) := -\frac{X(X-1)(X+1-2Z)}{2Z^3}.$$

Thus, given Z, the update is computed by first selecting the smallest admissible solution $X \in [1, Z)$ of this equation, and then setting $Z^+ = Z^2/X^2$. If no such solution exists, the transition fails.

Fixed points of this scalar map characterize self-sustained retrieval across repeated transitions: $Z^+ = Z$ means that each jump regenerates the same saliency amplitude. Since $Z^+ = Z^2/X^2$, fixed points satisfy $Z = X^2$, and the implicit equation for X becomes

$$\frac{1}{\kappa} = G(X^2, X) = \frac{(X-1)^2(2X+1)}{2X^5}.$$

Thus fixed points exist when the level $1/\kappa$ intersects the curve G. Since G has a unique maximum at $\bar{X} = (5+\sqrt{105})/8$, with $G(\bar{X}^2, \bar{X}) \simeq 0.0785$, fixed points exist only for $\kappa \geq \kappa_{\text{crit}} := 1/G(\bar{X}^2, \bar{X}) \simeq 12.74$. In particular, for $\kappa > \kappa_{\text{crit}}$, two fixed points appear $\bar{Z}_1$ and $\bar{Z}_2$, $\bar{Z}_1 < \bar{Z}_2$, with $\bar{Z}_1$ unstable and $\bar{Z}_2$ stable. Since $X = e^{T_{\text{esc}}}$ and $\bar{Z}_2 = X^2$ at the stable fixed point, the asymptotic escape time is $T^\infty_{\text{esc}} = \frac{1}{2}\log \bar{Z}_2$. Further details are given in Appendix D.

Fig. 4 illustrates the preceding analysis and extends it numerically to the case $\psi = \tanh$. Fig. 4(B) compares the SmoothF and Tanh activation functions, while Fig. 4(A) shows the induced saliency

[4]The range constraint $X \in [1, Z/\phi^{-1}(1))$ follows from $T_{\text{esc}} \in [0, T_{\max})$. In fact, since $z_\nu(t) = Ze^{-t}$ and $\phi(Ze^{-T_{\max}}) = 1$, we have $e^{T_{\max}} = Z/\phi^{-1}(1)$.

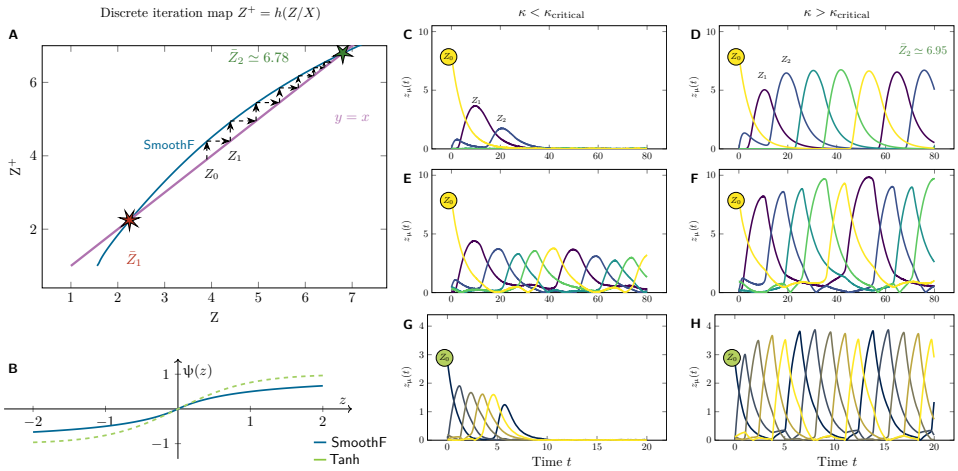

Figure 4: (A) Fixed points of the discrete map $Z^+ = h(Z/X)$ with $\kappa = 15$ for the SmoothF activation, with predicted $\bar{Z}_2 \simeq 6.78$ (see Fig. 9 in Appendix E for $\psi = \tanh$). Real solutions $\bar{Z}_1, \bar{Z}_2 > 1$ emerge only for $\kappa \geq \kappa_{\text{crit}} \simeq 12.74$; self-sustained transitions require $Z_0 > \bar{Z}_1$. (B) SmoothF (blue) and Tanh (green) activation functions. (C, D) Orthogonal memories with SmoothF activation. (C) Subcritical gain ($\kappa = 9$): oscillations decay and retrieval collapses to the origin. (D) Supercritical gain ($\kappa = 15$): persistent oscillations converge to the stable fixed point $\bar{Z}_2 \simeq 6.95$, close to theoretical prediction, enabling reliable sequential retrieval. (E, F) Random memories with SmoothF activation. (E) Subcritical gain: cross-talk noise sustains small-amplitude transitions. (F) Supercritical gain: transitions persist but exhibit amplitude fluctuations without entrainment to $\bar{Z}_2$. (G,H) Random memories with Tanh activation. (G) Subcritical gain ($\kappa = 3$) leads to collapse, with slower decay due to cross-talk. (H) Supercritical gain ($\kappa = 5$) yields self-sustained transitions. Panels (C-H) use $N = 256$, $P = 15$, and a directed cycle of length $L = 5$.

update map for SmoothF, revealing the emergence of two fixed points for gains exceeding the critical value $\kappa > \kappa_{\text{crit}} \simeq 12.74$. Fig. 4(C)-(F) present representative trajectories for subcritical and supercritical values of κ using the SmoothF activation. For orthogonal memories (Fig. 4(C)-(D)), subcritical gains lead to collapse into an inactive state, whereas supercritical gains give rise to self-sustained sequential retrieval. In contrast, for random memories (Fig. 4(E)-(F)), both gain regimes exhibit sustained transitions, likely due to cross-talk noise and the consequent lack of entrainment to a fixed asymptotic value. In the subcritical regime, the resulting z-oscillations have small amplitude and may therefore lead to collapse only on longer timescales. Finally, Fig. 4(G)-(H) show that the qualitative features of the slow z-dynamics persist persist with Tanh activation and random memories.

We next validate the theoretical escape-time predictions through simulations of the full model. For orthogonal memories, numerical escape times closely match the predictions (Fig. 5(A)), while the switching times T_{switch}, defined as the times at which the network transitions to the successor memory, occur slightly later due to the finite time required to leave the unstable memory state. Agreement between theoretical and numerical asymptotic escape times $T_{\text{esc}}^{\infty} = \frac{1}{2} \log \bar{Z}_2$ improves with κ (Fig. 5(B)). Adding Gaussian noise to the fast dynamics reduces the mean switching delay $T_{\text{switch}} - T_{\text{esc}}$, eventually making it negative due to premature switching (Fig. 5(C)).

We then investigate robustness to non-orthogonal memories by independently sampling biased binary patterns, as in [19], according to $\mathbb{P}(\xi_i^{\mu} = \pm 1) = (1 \pm \epsilon)/2$, where $\epsilon \in [0, 1]$ induces an expected normalized pairwise overlap of ϵ^2. Fig. 5(D) shows that the critical gain for successful retrieval increases only slightly with memory bias ϵ, while retrieval reliability decreases sharply for $\epsilon \gtrsim 0.4$. No successful transitions are observed for $\epsilon > 0.5$ over the investigated gains. Conditional on successful transitions, both the asymptotic escape time (Fig. 5(E)) and saliency amplitude (Fig. 5(F)) increase with κ. As memory bias ϵ increases, escape times decrease, while saliency remains approximately constant. Overall, these results suggest that the theoretical estimates derived for orthogonal memories remain informative for moderately correlated memories ($\epsilon \lesssim 0.35$), while stronger correlations lead to significant deviations and eventually compromise sequential retrieval. Additional numerical results for the Tanh activation are provided in Appendix E.

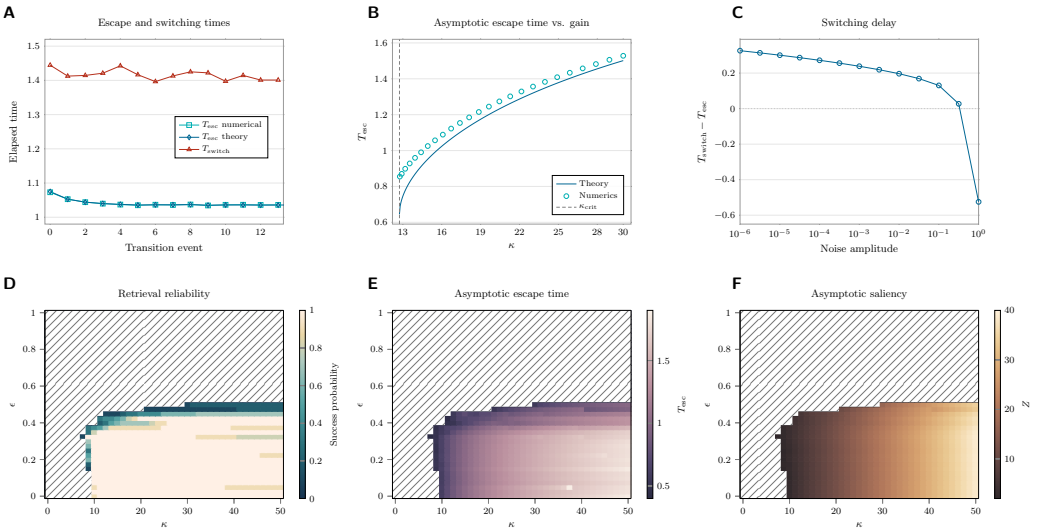

Figure 5: (A) Theoretical and numerical escape times $T_{\rm esc}$ and observed switching times $T_{\rm switch}$ across consecutive transitions ($\kappa = 15$, $Z_0 = 8$). (B) Asymptotic escape time versus κ: theory and numerical estimates. (C) Mean switching delay $T_{\rm switch} - T_{\rm esc}$ versus Gaussian noise amplitude σ. (D) Probability of correct transitions in the (κ, ϵ) plane. (E) Asymptotic escape time versus κ and memory bias ϵ. (F) Asymptotic saliency amplitude versus κ and ϵ. The numerical configuration is the same as in Fig. 4. Panels (A-C) use orthogonal memories, while (D-F) use biased random memories, with 10 independent realizations per parameter pair. In (B), asymptotic escape times are estimated over the last 8 time units of simulations of duration $T = 40$ (approximately 5-7 transitions). Panel (C) uses Gaussian noise with standard deviation $\sigma/\sqrt{N}$ and 64 realizations per noise amplitude. Hatched regions in (D-F) indicate no successful transitions.

6 Limitations and future research

While this work presents an analytically tractable framework for sequential retrieval in Hopfield networks, several important limitations remain. First, our analysis is restricted to canonical Tanh-type activation functions and does not address the practically relevant case $\Psi = \mathrm{softmax}$. This regime is of particular interest, as it directly links Hopfield dynamics to self-attention mechanisms in Transformer architectures [10] and is known to enable exponential memory capacity [9]. Extending the present theory to non-diagonal activation functions represents a promising direction for future work, with the potential to provide dynamical interpretations of inference in modern Transformer-based models. Second, although we provide a qualitative geometric characterization of the invariant sets and unstable structures supporting sequential retrieval, we do not yet precisely quantify the ranges of slow-fast timescale separation for which transitions emerge intrinsically, without reliance on numerical error or small-amplitude noise to trigger adherence to unstable fibers. A more refined geometric analysis is needed to delineate the parameter regimes that permit robust, noise-free sequential transitions. Third, while our results establish conditions for sequential retrieval, they do not characterize how retrieval reliability scales with the number of stored memories and the length of the retrieved sequence. Quantifying these capacity limits remains an important open problem, preliminarily explored numerically in Appendix F. Fourth, our theoretical analysis is restricted to deterministic transition graphs with at most one outgoing edge per memory. Extending the theory to branching graphs is another promising direction, as preliminary simulations in Appendix G suggest that memory overlap can influence branch selection. Finally, the role of external inputs in controlling retrieval sequences remains largely unexplored. Developing a theoretical framework for steering memory transitions through external signals is an interesting direction for future research, illustrated by the associative-memory finite-state machine in Appendix H.

References

[1] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, April 1982. ISSN 1091-6490. doi: 10.1073/pnas.79.8.2554.

[2] M. A. Cohen and S. Grossberg. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(5): 815–826, September 1983. ISSN 2168-2909. doi: 10.1109/tsmc.1983.6313075.

[3] J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 81(10):3088–3092, May 1984. ISSN 1091-6490. doi: 10.1073/pnas.81.10.3088.

[4] D. J. Amit, H. Gutfreund, and H. Sompolinsky. Information storage in neural networks with low levels of activity. *Physical Review A*, 35(5):2293–2303, March 1987. ISSN 0556-2791. doi: 10.1103/physreva.35.2293.

[5] R. McEliece, E. Posner, E. Rodemich, and S. Venkatesh. The capacity of the hopfield associative memory. *IEEE Transactions on Information Theory*, 33(4):461–482, July 1987. ISSN 1557-9654. doi: 10.1109/tit.1987.1057328.

[6] M. V. Tsodyks and M. V. Feigel'man. The enhanced storage capacity in neural networks with low activity level. *Europhysics Letters (EPL)*, 6(2):101–105, May 1988. ISSN 1286-4854. doi: 10.1209/0295-5075/6/2/002.

[7] D. Krotov and J. J. Hopfield. Dense associative memory for pattern recognition. In *Advances in neural information processing systems*, volume 29, 2016. doi: doi.org/10.48550/arXiv.1606.01164.

[8] D. Krotov and J. J. Hopfield. Large associative memory problem in neurobiology and machine learning. In *International Conference on Learning Representations*, 2020. doi: doi.org/10.48550/arXiv.2008.06996.

[9] M. Demircigil, J. Heusel, M. Lowe, S. Upgang, and F. Vermet. On a model of associative memory with huge storage capacity. *Journal of Statistical Physics*, 168(2):288–299, May 2017. ISSN 1572-9613. doi: 10.1007/s10955-017-1806-y.

[10] H. Ramsauer, B. Schäfl, J. Lehner, P. Seidl, M. Widrich, L. Gruber, M. Holzleitner, T. Adler, D. Kreil, M. K. Kopp, G. Klambauer, J. Brandstetter, and S. Hochreiter. Hopfield networks is all you need. In *International Conference on Learning Representations*, 2021.

[11] B. Hoover, Y. Liang, B. Pham, R. Panda, H. Strobelt, D. H. Chau, M. Zaki, and D. Krotov. Energy transformer. *Advances in neural information processing systems*, 36:27532–27559, 2023.

[12] B. Hoover, H. Strobelt, D. Krotov, J. Hoffman, Z. Kira, and D. H. Chau. Memory in plain sight: A survey of the uncanny resemblances between diffusion models and associative memories. In *Associative Memory & Hopfield Networks in 2023*, 2023.

[13] L. Ambrogioni. In search of dispersed memories: Generative diffusion models are associative memory networks. *Entropy*, 26(5):381, April 2024. ISSN 1099-4300. doi: 10.3390/e26050381.

[14] D. Kleinfeld. Sequential state generation by model neural networks. *Proceedings of the National Academy of Sciences*, 83(24):9469–9473, December 1986. ISSN 1091-6490. doi: 10.1073/pnas.83.24.9469.

[15] H. Sompolinsky and I. Kanter. Temporal association in asymmetric neural networks. *Physical Review Letters*, 57(22):2861–2864, December 1986. ISSN 0031-9007. doi: 10.1103/physrevlett.57.2861.

[16] D. W. Tank and J. J. Hopfield. Neural computation by concentrating information in time. *Proceedings of the National Academy of Sciences*, 84(7):1896–1900, April 1987. ISSN 1091-6490. doi: 10.1073/pnas.84.7.1896.

[17] D. Kleinfeld and H. Sompolinsky. Associative neural network model for the generation of temporal patterns. theory and application to central pattern generators. *Biophysical Journal*, 54(6):1039–1051, December 1988. ISSN 0006-3495. doi: 10.1016/s0006-3495(88)83041-8.

[18] A. Karuvally, T. Sejnowski, and H. T. Siegelmann. General sequential episodic memory model. In *International Conference on Machine Learning*, pages 15900–15910. PMLR, 2023.

[19] H. T. Chaudhry, J. A. Zavatone-Veth, D. Krotov, and C. Pehlevan. Long sequence hopfield memory. *Journal of Statistical Mechanics: Theory and Experiment*, 2024(10):104024, October 2024. ISSN 1742-5468. doi: 10.1088/1742-5468/ad6427.

[20] A. Karuvally, P. Lertsaroj, S. Terrence, and H. T. Siegelmann. Exponential dynamic energy network for high capacity sequence memory. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. doi: 10.48550/arXiv.2510.24965.

[21] L. Herron, P. Sartori, and B. Xue. Robust retrieval of dynamic sequences through interaction modulation. *PRX Life*, 1(2), December 2023. ISSN 2835-8279. doi: 10.1103/prxlife.1.023012.

[22] M. Farrell and C. Pehlevan. Recall tempo of hebbian sequences depends on the interplay of hebbian kernel with tutor signal timing. *Proceedings of the National Academy of Sciences*, 121(32), 2024. ISSN 1091-6490. doi: 10.1073/pnas.2309876121.

[23] S. Betteti, G. Baggio, F. Bullo, and S. Zampieri. Input-driven dynamics for robust memory retrieval in hopfield networks. *Science Advances*, 11(17), April 2025. ISSN 2375-2548. doi: 10.1126/sciadv.adu6991.

[24] M. K. Benna and S. Fusi. Computational principles of synaptic memory consolidation. *Nature Neuroscience*, 19(12):1697–1706, October 2016. ISSN 1546-1726. doi: 10.1038/nn.4401.

[25] M. Gillett and N. Brunel. Dynamic control of sequential retrieval speed in networks with heterogeneous learning rules. *eLife*, 12, August 2024. ISSN 2050-084X. doi: 10.7554/elife.88805.3.

[26] N. Fenichel. Geometric singular perturbation theory for ordinary differential equations. *Journal of Differential Equations*, 31(1):53–98, January 1979. ISSN 0022-0396. doi: 10.1016/0022-0396(79)90152-9.

[27] G. Hek. Geometric singular perturbation theory in biological practice. *Journal of Mathematical Biology*, 60(3):347–386, April 2009. ISSN 1432-1416. doi: 10.1007/s00285-009-0266-7.

[28] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, December 1985. ISBN 9780511810817.

Appendix

A The IDP Hopfield model

The IDP Hopfield model presented in [23] is characterized by the O.D.E.

$$\dot{x} = -x + W(\alpha)\Psi(x) \qquad \alpha \in \mathbb{R}^P_{>0} \tag{15}$$

where $W(\alpha)$ is the synaptic matrix of neural interaction built from the memories of the model. The memories $\xi^1, \dots, \xi^P$ and associated synaptic matrix $\mathrm{W}(\alpha)$ obey the following assumption.

Assumption 1 (Memory and synaptic matrix). *Let $P \in \mathbb{N}$ and define the set of orthogonal memories*

$$\Sigma := \{\xi^1, \dots, \xi^P\} \subseteq \{-1, 1\}^N \tag{16}$$

$$\xi^{\mu\top}\xi^{\nu} = N\delta_{\mu,\nu} \qquad \mu, \nu \in \{1, \dots, P\} \tag{17}$$

where $\delta_{\mu,\nu}$ is the Kronecker delta for indices μ and ν. For $\alpha \in \mathbb{R}^P_{>0}$ positive we define the input-modulated synaptic matrix $\mathrm{W}(\alpha) \in \mathbb{R}^{N\times N}$ as

$$\mathrm{W}(\alpha) = \frac{1}{N}\sum_{\mu=1}^{P} \alpha_\mu \xi^\mu \xi^{\mu\top}. \tag{18}$$

The orthogonal memories constraint is necessary to derive rigorous results on existence and stability of fixed points, but can be relaxed to orthogonality in expectation in implementations:

$$\mathbb{P}[\xi_i^\nu = 1] = \mathbb{P}[\xi_i^\nu = -1] = \tfrac{1}{2} \qquad \forall \nu = 1, \dots, P, \quad \forall i = 1, \dots, N. \tag{19}$$

Remark 1. *Notice that by constructing the memory matrix*

$$M = \tfrac{1}{\sqrt{N}}[\xi^1, \dots, \xi^P] \tag{20}$$

we can rewrite the input-modulated synaptic matrix as

$$W(\alpha) = M\mathrm{diag}(\alpha)M^\top \tag{21}$$

The canonical synaptic matrix W of the Hopfield model is a specific instance of (18) for the choice $\alpha = \mathbb{1}$ (or a scalar multiple of it).

We now introduce an assumption on the saliency function Φ going from the space of the slow neural feedback $z \in \mathbb{R}^P$ to the saliency space $\alpha \in \mathbb{R}^P$ such that $\alpha = \Phi(z)$.

Assumption 2. *The saliency function $\Phi : \mathbb{R}^P \to \mathbb{R}^P$ satisfies the following properties:*

(i) ***(diagonality and homogeneity)*** *for all $\nu = 1, \dots, P$ we have $\Phi_\nu(z) = \phi(z_\nu)$;*

(ii) ***(symmetry and positivity)*** *$\phi(s) = \phi(-s) \geq 0$, with $\alpha_\nu(0) = 0$;*

(iii) ***(symmetric monotonicity)*** *$\phi(s)$ is monotonically increasing on $(0, +\infty)$.*

Likewise, we introduce some smoothness assumptions on the activation function Ψ of the IDP Hopfield network, which will be useful for the characterization of the existence and local stability conditions for the equilibria of (15).

Assumption 3. *The activation function $\Psi : \mathbb{R}^N \to \mathbb{R}^N$ satisfies the following properties:*

(i) ***(diagonality, homogeneity and smoothness)*** *for all $i = 1, \dots, N$, we have $\Psi_i(x) = \psi(x_i)$ with $\psi \in C^1(\mathbb{R})$;*

(ii) ***(odd and saturating)*** *$\psi(s) = -\psi(-s)$ for all $s \in \mathbb{R}$ and $\lim_{s\to\pm\infty} \psi(s) = \pm 1$.*

(iii) ***(bounded monotonicity)*** *$\psi'(s) \in (0, 1]$ for all $s \in \mathbb{R}$ and $\psi'(0) = 1$;*

(iv) ***(monotonicity of the derivative)*** *ψ' is monotonically decreasing on $(0, +\infty)$.*

We will now state some of the main analytical results presented in [23, Supplementary Information], starting from the existence conditions for the equilibria of (15). The presented results are stated with respect to the saliency states $\alpha = \Phi(z)$.

Theorem 2 (Existence of retrievable memories)**.** *Let ξ^ρ be a memory as in Assumption 1. Then $x_\rho = \gamma_\rho \xi^\rho$ is an equilibrium for (15) for some unique $\gamma_\rho \in \mathbb{R} \setminus \{0\}$ if and only if $\alpha_\rho > 1$. In this case, γ_ρ satisfies*

$$\alpha_\rho = f(\gamma_\rho) = \tfrac{\gamma_\rho}{\psi(\gamma_\rho)} \tag{22}$$

We now provide the main result on the local stability of the equilibria for the fast IDP Hopfield dynamics.

Theorem 3 (Local stability of equilibria and memories)**.** *Consider the dynamics (15). Then*

1. *an equilibrium $x^\star \in \mathbb{R}^N$ is locally exponentially stable if and only if*

$$-[D\Psi(x^\star)]^{-1} + W(\alpha) \prec 0 \tag{23}$$

2. *if $x_\rho = \gamma_\rho \xi^\rho$ with $\gamma_\rho \in \mathbb{R} \setminus \{0\}$ is an equilibrium, then x_ρ is locally exponentially stable if and only if*

$$\psi'(\gamma_\rho) < \frac{1}{\max_{\mu=1,\dots,P} \alpha_\mu} \tag{24}$$

In particular, the study of the local stability condition allows to provide exact conditions on the points where the Jacobian becomes has eigenvalues with null real part, which will later be useful for the characterization of the memory manifold (see Section C). We will restrict ourselves to equilibrium conditions characterized by $\gamma_\rho > 0$, as the case $\gamma_\rho < 0$ is associated to the so called anti-memory $-\xi^\rho$ and is a symmetric artifact of classical Hopfield models. Notice in particular that by property 3(iv) the derivative ψ' is locally invertible in $(0, +\infty)$, and we can define the following stability function.

Definition 1 (Saliency stability function)**.** *We define the saliency stability function $\omega : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ as*

$$\omega(s) = f \circ \psi'^{-1}\left(\tfrac{1}{s}\right) = \frac{\psi'^{-1}\left(\frac{1}{s}\right)}{\psi\left(\psi'^{-1}\left(\frac{1}{s}\right)\right)} \tag{25}$$

Before providing the key results on the eigenvalues of the Jacobian of the IDP Hopfield vector field, we establish the basic invertibility and monotonicity properties of the saliency stability function.

Lemma 1 (Invertibility of the saliency stability function)**.** *The saliency stability function $\omega : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is invertible, and in particular*

$$\omega^{-1}(r) = \frac{1}{\psi' \circ f^{-1}(r)}, \qquad r > 0. \tag{26}$$

Furthermore, it holds that

1. $\omega(s) < s$;
2. $\omega^{-1}(r) > r$.

Proof. The paragraph above Definition 1 establishes the invertibility of ψ' in $(0, +\infty)$. Considering then the function $f(\gamma) = \frac{\gamma}{\psi(\gamma)}$ we immediately observe that for any $\gamma > 0$ we have

$$f'(\gamma) = \frac{\psi(\gamma) - \gamma\psi'(\gamma)}{\psi'(\gamma)^2}. \tag{27}$$

From Assumption 3(iv) we have that ψ' is monotonically decreasing on $(0, +\infty)$ and in particular we have

$$\psi(\gamma) = \int_0^\gamma \psi'(s)\, ds > \int_0^\gamma \psi'(\gamma)\, ds = \gamma\psi'(\gamma). \tag{28}$$

Consequently $f'(\gamma) > 0$ for all $\gamma > 0$ and in particular f is invertible on $(0, +\infty)$. Taking $\omega^{-1}(r) = 1/(\psi' \circ f^{-1}(r))$ we observe that

$$\omega^{-1} \circ \omega(s) = \frac{1}{\psi' \circ f^{-1} \circ f \circ \psi'^{-1}\left(\frac{1}{s}\right)} \tag{29}$$

$$= \frac{1}{s^{-1}} = s \tag{30}$$

The proof of points 1-2 is straightforward, and relies on the fact that $\gamma\psi'(\gamma) - \psi(\gamma) < 0$ for all $\gamma > 0$, as proven in (28).

1. $\omega(s) < s$:
   We are interested in proving that
   $$\omega(s) - s = f \circ \psi'^{-1}\left(\tfrac{1}{s}\right) - s < 0. \tag{31}$$
   Taking $\gamma = \psi'^{-1}\left(\frac{1}{s}\right)$, we can express $s = \frac{1}{\psi'(\gamma)}$, and therefore our condition reduces to
   $$f(\gamma) - \frac{1}{\psi'(\gamma)} = \frac{\gamma}{\psi(\gamma)} - \frac{1}{\psi'(\gamma)} < 0. \tag{32}$$

2. $\omega^{-1}(r) > r$:
   We are interested in proving that
   $$\tfrac{1}{\omega^{-1}(r)} - \tfrac{1}{r} = \psi' \circ f^{-1}(r) - \tfrac{1}{r} < 0. \tag{33}$$
   Taking $\gamma = f^{-1}(r)$, we can express $r = f(\gamma)$, and therefore our condition reduces to
   $$\psi'(\gamma) - \frac{1}{f(\gamma)} = \psi'(\gamma) - \frac{\psi(\gamma)}{\gamma} < 0. \tag{34}$$

□

The saliency stability function allows us to state an important proposition on the eigenvalues of the Jacobian of 15 at the equilibrium points.

Proposition 1 (Eigenvalues of the IDP Hopfield Jacobian)**.** *Let $\alpha \in \mathbb{R}^P_{>0}$ and define $F(x) = -x + W(\alpha)\Psi(x)$. Let $\alpha_\nu > 1$ such that for $\gamma_\nu = f^{-1}(\alpha_\nu)$ the equilibrium $x_\nu = \gamma_\nu \xi^\nu$ exists. Then the Jacobian matrix*

$$\frac{\partial F}{\partial x}(x_\nu) = -\boldsymbol{I} + W(\alpha)\frac{\partial \Psi}{\partial x}(x_\nu) \tag{35}$$

has one eigenvalue λ_μ with $\Re\{\lambda_\mu\} = 0$ iff

$$\exists \mu \in \{1, \dots, P\} : \qquad \alpha_\nu = \omega(\alpha_\mu). \tag{36}$$

Proof. From [23, Theorems 2, SI] we have that

$$\frac{\partial F}{\partial x}(x_\nu) = -\boldsymbol{I} + W(\alpha)\frac{\partial \Psi}{\partial x}(x_\nu) \tag{37}$$

This has the same spectrum of the matrix

$$\frac{\partial \Psi}{\partial x}(x_\nu)^{1/2}\frac{\partial F}{\partial x}(x_\nu)\frac{\partial \Psi}{\partial x}(x_\nu)^{-1/2} = \frac{\partial \Psi}{\partial x}(x_\nu)^{1/2}\left(-\frac{\partial \Psi}{\partial x}(x_\nu)^{-1} + W(\alpha)\right)\frac{\partial \Psi}{\partial x}(x_\nu)^{1/2}$$

This matrix has the same inertia [28] of the matrix

$$DF(x_\nu) = -\frac{\partial \Psi}{\partial x}(x_\nu)^{-1} + W(\alpha)$$

that has spectrum

$$\rho\left(DF(x_\nu)\right) := \Big\{\alpha_1 - \frac{1}{\psi'(\gamma_\nu)}, \dots, \alpha_P - \frac{1}{\psi'(\gamma_\nu)}, \underbrace{-1, \dots, -1}_{N-P}\Big\}, \tag{38}$$

From this it is immediate to see that $\frac{\partial F}{\partial x}(x_\nu)$ has null eigenvalues iff there exists $\mu \in \{1, \dots, P\}$ such that

$$\psi'(\gamma_\nu) = \frac{1}{\alpha_\mu}. \tag{39}$$

Exploiting the invertibility of ψ' in $(0, +\infty)$, we get the equivalent condition

$$\gamma_\nu = \psi'^{-1}\left(\tfrac{1}{\alpha_\mu}\right) \quad \exists \mu \in \{1, \dots, P\}. \tag{40}$$

Since $\alpha_\nu > 1$, then exploiting the uniqueness of $\gamma_\nu > 0$ in the equilibrium condition in (22) we have that the zero eigenvalue condition in (40) coincides with

$$\begin{aligned} \alpha_\nu &= f(\gamma_\nu) \\ &= f \circ \psi'^{-1}\left(\tfrac{1}{\alpha_\mu}\right) = \omega(\alpha_\mu) \quad \exists \mu \in \{1, \dots, P\}. \end{aligned} \tag{41}$$

□

In particular, notice that from Lemma 1 for all $\nu = 1, \dots, P$ it holds that $\alpha_\nu > \omega(\alpha_\nu)$, so that whenever at least a memory exists, at least one is also locally exponentially stable. Notice in addition that the number of eigenvalues with positive and negative real part is obtained by counting the instances where the condition $\alpha_\nu \lessgtr \omega(\alpha_\mu)$ is satisfied for $\mu = 1, \dots, P$.

B A modern reformulation of the sequential IDP Hopfield model

The IDP Hopfield model admits a modern Hopfield [8] interpretation in which retrieval is decomposed into three interacting processes: a memory layer that reads out the similarity between the current feature state and the stored memories, a saliency layer that modulates this readout, and a feature layer that reconstructs the retrieved representation [23]. In the original IDP model, the saliency layer is driven by the external input. Here we show that the sequential IDP Hopfield model (3) admits the same interpretation, with the saliency layer driven instead by internal graph-structured feedback.

We introduce an auxiliary memory-index variable $y \in \mathbb{R}^P$ and consider

$$\begin{cases} \tau_x \dot{x} = -x + M \mathrm{diag}(\Phi(z)) y, \\ \tau_y \dot{y} = -y + M^\top \Psi(x), \\ \tau_z \dot{z} = -z + \chi(y). \end{cases}$$

The variable y is the memory-layer activity: its components encode the alignment between the current activated feature state and the stored memories. The map $\chi : \mathbb{R}^P \to \mathbb{R}^P$ is a slow feedback map prescribing the target saliency profile toward which z relaxes.

Under the timescale separation $\tau_y \ll \tau_x \ll \tau_z$, the memory layer rapidly satisfies

$$y \simeq M^\top \Psi(x).$$

Substituting this quasi-steady relation into the feature dynamics gives

$$\tau_x \dot{x} = -x + M \mathrm{diag}(\Phi(z)) M^\top \Psi(x),$$

which is exactly the fast equation of the sequential IDP Hopfield model (3).

The slow equation in (3) is recovered by choosing

$$\chi(y) = \frac{\kappa}{\sqrt{N}} A y.$$

Indeed, replacing y by its fast quasi-steady value yields

$$\tau_z \dot{z} = -z + \frac{\kappa}{\sqrt{N}} A M^\top \Psi(x),$$

which coincides with the slow feedback dynamics of (3).

This reformulation clarifies the relation with the modern IDP interpretation. In the original IDP model, saliencies are prescribed by an external input decomposition. In the sequential IDP model, saliencies are generated autonomously by the retrieved memory through the transition graph.

C Memory manifolds: pouring the concrete

In the classic slow-fast characterization from geometric singular perturbation theory (GSPT) presented in [26], the timescale separation between coupled dynamics is extremized in order to derive geometric information about the loci in state-space where the system evolves. In the case of the IDP Hopfield model, the coupled system can be written as

$$\begin{cases} \varepsilon \dot{x} = -x + M \operatorname{diag}(\phi(z))\, M^\top \Psi(x), & x(0) \in \mathbb{R}^N, \\ \dot{z} = -z + \dfrac{\kappa}{\sqrt{N}}\, A M^\top \Psi(x), & z(0) \in \mathbb{R}^P. \end{cases} \tag{42}$$

and taking the limit $\varepsilon \to 0^+$, the problem simplifies to studying how the dynamics

$$\dot{z} = -z + \frac{\kappa}{\sqrt{N}} A M^\top \Psi(x^\star(z)) \tag{43}$$

evolve under the solution constraint, parametrized as a function of the slow variable z, $\mathbb{0}_N = -x^\star(z) + M\mathrm{diag}(\Phi(z))M^\top \Psi(x^\star(z))$. We start by studying the solutions of the algebraic constraint, which characterize the critical set

$$\mathcal{S}_0 := \{(x,z) \in \mathbb{R}^N \times \mathbb{R}^P :\ x = M\mathrm{diag}(\Phi(z))M^\top \Psi(x)\} \tag{44}$$

C.1 Memory branches and memory manifolds

Exploiting Theorem 2 for the existence of equilibria in the IDP Hopfield model, we proceed to define the memory branches, which are disjoint sets in $\mathcal{S}_0$ associated each to a different memory.

Definition 2. *Let $\phi(z_\nu) > 1$. Then the νth memory branch is the P-dimensional set $\mathcal{N}_\nu \subseteq \mathcal{S}_0$ defined as*

$$\mathcal{N}_\nu = \left\{(x,z) \in \mathbb{R}^N \times \mathbb{R}^P :\ \exists \gamma_\nu \neq 0,\ x = \gamma_\nu \xi^\nu,\ \phi(z_\nu) = f(\gamma_\nu)\right\}, \tag{45}$$

Notice that the memory branch $\mathcal{N}_\nu$ is not defined for $\phi(z_\nu) < 1$, as the equilibria associated to the memory ξ^ν does not exists. The memory manifold is then defined as a P-dimensional manifold comprising the points in $\mathcal{N}_\nu$ where normal hyperbolicity holds.

Definition 3. *Let $F(x,z) = -x + W(\Phi(z))\Psi(x)$ and assume $\phi(z_\nu) > 1$. Then the νth memory manifold $\mathcal{M}_\nu$ is the set of points $(x,z) \in \mathcal{N}_\nu$ such that*

$$\nexists \lambda \in \rho\left(\frac{\partial F}{\partial x}(x,z)\right) :\quad \Re(\lambda) = 0. \tag{46}$$

Of particular relevance are two disjoint regions of the manifold $\mathcal{M}_\nu$, which play qualitatively distinct roles in shaping the system's long-term behavior. Rather than pursuing formal definitions, we give an intuitive description aimed at conveying their dynamical function:

- *Attracting manifold* $\mathcal{M}_\nu^a$: this region consists of points that act as local sinks for the dynamics. Trajectories initialized in a neighborhood of these points in the ambient space are drawn toward the manifold and remain confined to it over time. Geometrically, $\mathcal{M}_\nu^a$ behaves as a stabilizing surface that captures and retains nearby trajectories.
- *Saddle manifold* $\mathcal{M}_\nu^s$: this region contains points that are simultaneously attracting and repelling, depending on direction. Trajectories may approach the manifold transiently, but the presence of unstable directions eventually pushes them away. As a result, $\mathcal{M}_\nu^s$ acts as a transient conduit, enabling trajectories to leave one attracting region and move toward another.

One of the advantages of relying on the IDP Hopfield model to capture sequentiality and memory transition is that there already exists a complete theory on the stability of system's equilibria. We exploit exactly this theory to characterize the attracting and saddle memory manifolds.

Lemma 2 (Attracting and saddle manifolds)**.** *Let $\omega : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ be the saliency stability function in Definition 1. Then the point $(x,z) \in \mathcal{M}_\nu$ on the memory manifold is*

- *attracting if $\phi(z_\nu) > \omega(\phi(z_\mu))$ for all $\mu \in \{1,\dots,P\}$;*

- *saddle type if there exists at least one* $\mu \in \{1, \dots, P\}/\{\nu\}$ *such that* $\phi(z_\nu) < \omega(\phi(z_\mu))$.

Proof. The manifold $\mathcal{M}_\nu$ is attracting at (x, z) if for all

$$\lambda \in \rho\left(\frac{\partial F}{\partial x}(x, z)\right) \quad \Re(\lambda) < 0 \tag{47}$$

and a saddle if there exists at least an eigenvalue λ such that $\Re(\lambda) > 0$. From Proposition 1, we have that the eigenvalues of the Jacobian $\frac{\partial F}{\partial x}$ at $(x, z) \in \mathcal{M}_\nu$ have the same sign of $\{\phi(z_\mu) - \frac{1}{\psi'(\gamma_\nu)}\}_{\mu=1}^P$ with $\gamma_\nu = f^{-1}(\phi(z_\nu))$. Thus, $\mathcal{M}_\nu$ is attracting at (x, z) iff for all $\mu = 1, \dots, P$

$$\phi(z_\mu) - \frac{1}{\psi'(\gamma_\nu)} < 0 \quad \iff \quad \phi(z_\nu) > \omega(\phi(z_\mu)). \tag{48}$$

Conversely, it is a saddle if there exists $\mu \in \{1, \dots, P\}/\{\nu\}$ such that

$$\phi(z_\mu) - \frac{1}{\psi'(z_\nu)} > 0 \quad \iff \quad \phi(z_\nu) < \omega(\phi(z_\mu)). \tag{49}$$

□

The phenomena known as retrieval and fixation to the memory ξ^ν occur for points $(x, z) \in \mathcal{M}_\nu$ such that the memory manifold is attracting, i.e., the respective saliency weight $\phi(z_\nu)$ is greater than the saliency stability function computed at any $\phi(z_\mu)$. Escape direction emerge instead when the dynamics evolve towards a saddle region of the memory manifold. In general, we have that the attracting manifold for the memory ξ^ν is the subset $\mathcal{M}_\nu^a := \{(x, z) \in \mathcal{M}_\nu : \ \phi(z_\nu) > \omega(\phi(z_\mu)) \quad \forall \mu = 1, \dots, P\}$. Conversely, the saddle manifold is $\mathcal{M}_\nu^s := \{(x, z) \in \mathcal{M}_\nu : \ \exists \mu = 1, \dots, P \quad \phi(z_\nu) < \omega(\phi(z_\mu))\}$.

C.2 The origin manifold

In our description of the geometric sets where the slow-fast activity of the system is constrained, we considered all the equilibria solutions of the fast IDP Hopfield dynamics that are associated to one of the memories. However, there is another solution that is independent from any of the memory vectors: the origin.

Definition 4 (Origin manifold)**.** *The origin manifold is defined as the set*

$$\mathcal{M}_0 := \left\{(\mathbb{0}_N, z) \in \mathbb{R}^N \times \mathbb{R}^P : \quad \nexists \lambda \in \rho\left(\frac{\partial F}{\partial x}(\mathbb{0}_N, z)\right), \ \Re\{\lambda\} = 0\right\}. \tag{50}$$

In particular, for any point in the set $\mathcal{M}_0$ *we have that*

$$\frac{\partial F}{\partial x}(\mathbb{0}_N, z) = -\boldsymbol{I} + M \mathrm{diag}(\Phi(z)) M^\top \tag{51}$$

Notice that the structure of the Jacobian $\frac{\partial F}{\partial x}(\mathbb{0}_N, z)$ allows for a clean characterization of the attracting and saddle regions. Noticing that the spectrum of Jacobian is

$$\rho\left(\tfrac{\partial F}{\partial x}(\mathbb{0}_N, z)\right) := \{-1 + \phi(z_1), \dots, -1 + \phi(z_P), \underbrace{-1, \dots, -1}_{N-P}\}, \tag{52}$$

we can characterize the not normally hyperbolic points as $(\mathbb{0}_N, z)$ such that there exists $\mu = 1, \dots, P$ for which $\phi(z_\mu) = 1$. Similarly, we can characterize the:

1. attracting origin manifold: $\mathcal{M}_0^a := \{(\mathbb{0}_N, z) : \quad \|\Phi(z)\|_\infty < 1\}$;
2. saddle origin manifold: $\mathcal{M}_0^s := \mathcal{M}_0 \setminus \mathcal{M}_0^a$.

Of particular relevance is the attracting origin manifold $\mathcal{M}_0^a$, since it is a sink for the slow-fast dynamics, and once trajectories enter this set they cannot escape it. Thereby, from $\mathcal{M}_0^a$ the activity inevitably collapses to the origin $(\mathbb{0}_N, \mathbb{0}_P) \in \mathbb{R}^N \times \mathbb{R}^P$, as formalized in the following Lemma.

Lemma 3 (Collapse to the origin)**.** *Let* $(x_0, z_0) \in \mathcal{M}_0^a$. *Then* $(x(t), z(t)) \to (\mathbb{0}_N, \mathbb{0}_P)$ *exponentially fast.*

Proof. For $(x_0, z_0) \in \mathcal{M}_0^a$ the slow dynamics reduce to

$$\begin{aligned}\dot{z} &= -z + \frac{\kappa}{\sqrt{N}} A M^\top \Psi(\mathbb{0}_N) \\ &= -z. \end{aligned} \tag{53}$$

Therefore, the slow dynamics are completely dissipative and $z(t) = z_0 e^{-t}$. □

It is easy to observe that, in particular, $\mathcal{M}_0^a$ is also forward-invariant for the reduced dynamics.

D Slow saliency dynamics: under the hood

We provide the technical details for Sec. 4. We first prove Theorem 1, and then analyze the explicit scalar map introduced in the example.

D.1 Proof of Theorem 1

We start with two auxiliary results. The first one is a corollary of Lemma 1.

Corollary 4. *For every z such that $\phi(z) > 1$, we have*

$$\psi'\big(f^{-1}(\phi(z))\big) < \frac{1}{\phi(z)}.$$

Proof. From Lemma 1 we know that

$$\omega^{-1}(r) > r. \tag{54}$$

Taking $r = \phi(z)$ and the definition $\omega^{-1}(r) = 1/(\psi' \circ f^{-1}(r))$ we conclude. □

In the two-memory reduction of Sec. 4, the source and successor saliencies satisfy

$$z_\nu(t) = Ze^{-t}, \qquad \dot{z}_\mu = -z_\mu + \kappa g(Ze^{-t}), \qquad g := \psi \circ f^{-1} \circ \phi.$$

Hence

$$z_\mu(t) = \kappa \int_0^t e^{-(t-\tau)} g(Ze^{-\tau})\, d\tau. \tag{55}$$

Let $T_{\max}$ be defined by $\phi(Ze^{-T_{\max}}) = 1$. The next lemma shows that, once the successor saliency catches up with the source saliency, the source branch must lose stability before the source memory disappears (see also Fig. 6, right plot).

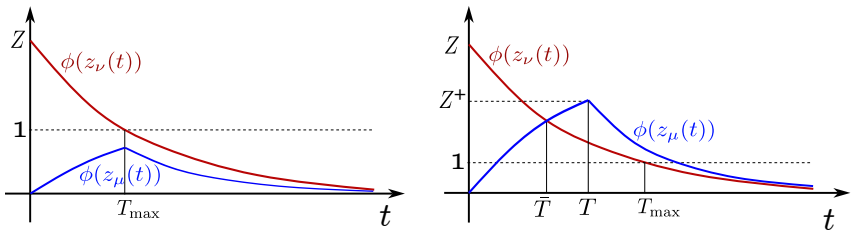

Figure 6: Graphical illustration of the two possible outcomes of the source-successor saliency dynamics. Left: no transition occurs and the dynamics decay to inactivity. Right: the successor saliency catches up with the source saliency, triggering a transition between memories.

Lemma 4. *Assume that there exists $\bar{T} \in [0, T_{\max})$ such that $z_\mu(\bar{T}) = z_\nu(\bar{T})$. Then there exists $T \in [\bar{T}, T_{\max})$ such that*

$$\omega^{-1}(\phi(z_\mu(T)) > \phi(z_\mu(T)) = \omega^{-1}(\phi(z_\nu(T)). \tag{56}$$

Proof. Observe first that the inequality on the left is always true by Corollary 4. We need to prove that the equality holds true for some $T \in [\bar{T}, T_{\max})$. To this aim we will prove that

$$\phi(z_\mu(t)) < \omega^{-1}(\phi(z_\nu(t)) \quad \text{for } t = \bar{T} \tag{57}$$

$$\phi(z_\mu(t)) > \omega^{-1}(\phi(z_\nu(t)) \quad \text{for } t = T_{\max}^- \tag{58}$$

which would prove that the equality should hold between $\bar{T}$ and $T_{\max}$.

Note that, again by Corollary 4, we have

$$\phi(z_\mu(\bar{T})) = \phi(z_\nu(\bar{T})) < \omega^{-1}(\phi(z_\nu(\bar{T}))$$

We want to prove now (58). To this aim let $\delta(t) := z_\mu(t) - z_\nu(t)$ and observe that $\delta(t)$ satisfies the same differential equation satisfied by $z_\mu(t)$ and hence we can argue that $\delta(t)$ is strictly increasing and $\delta(\bar{T}) = 0$. This implies that $\delta(T_{\max}^-) > 0$ and hence

$$\frac{1}{\phi(z_\mu(T_{\max}^-))} = \frac{1}{\phi(z_\nu(T_{\max}^-) + \delta(T_{\max}^-))} < \frac{1}{\phi(z_\nu(T_{\max}^-))} = 1$$

On the other hand observe that $\phi(z_\nu(T_{\max}^-)) = 1$ so that $f^{-1}(\phi(z_\nu(T_{\max}^-))) = 0$ and hence

$$\omega^{-1}(\phi(z_\nu(T_{\max}^-)) = 1$$

where we used the assumption $\psi'(0) = 1$. This yields the thesis. □

We now prove Theorem 1.

Proof of Theorem 1. By Lemma 4 and the IDP Hopfield stability condition in Theorem 3, if the successor saliency catches up with the source saliency before $T_{\max}$, then there exists a time in $[\bar{T}, T_{\max})$ at which the source branch loses stability. The first such time, T_{esc}, satisfies the identity in (56), namely

$$\phi(z_\mu(T_{\text{esc}})) = \omega^{-1}(\phi(z_\nu(T_{\text{esc}})).$$

Equivalently,

$$z_\mu(T_{\text{esc}}) = h\big(z_\nu(T_{\text{esc}})\big), \qquad h(z) := \phi^{-1}\big(\omega^{-1}(\phi(z))\big)\,.$$

Using $z_\nu(t) = Ze^{-t}$ and (55), and evaluating the previous identity at $t = T_{\text{esc}}$ yields

$$\kappa \int_0^{T_{\text{esc}}} e^{-(T_{\text{esc}}-\tau)} g(Ze^{-\tau})\, d\tau = h\big(Ze^{-T_{\text{esc}}}\big).$$

Thus T_{esc} is the smallest admissible solution of the previous equation, and the updated saliency is

$$Z^+ := z_\mu(T_{\text{esc}}) = h\big(Ze^{-T_{\text{esc}}}\big).$$

□

D.2 Extended derivations for the illustrative example

We give the details for the scalar iteration associated with

$$\psi(x) = \frac{x}{1+|x|}, \qquad \phi(z) = |z|.$$

In this case it can be proved that $f(x) = \frac{x}{\psi(x)} = x + 1$ for all $x \geq 0$ so that, for all $x \geq 1$ we have that $f^{-1}(x) = x - 1$. Moreover for $x \geq 0$ we have $\psi'(x) = \frac{1}{(x+1)^2}$. Thus, for $z \geq 1$,

$$g(z) = 1 - z^{-1} \qquad h(z) = z^2.$$

Consequently, for $Z > 1$ and $X \in [1, Z)$,

$$\int_1^X g\Big(\frac{Z}{s}\Big)\, ds = \int_1^X \Big(1 - \frac{s}{Z}\Big)\, ds = (X-1)\left(1 - \frac{X+1}{2Z}\right).$$

Substituting this expression into $R(Z,X)=0$ yields

$$\frac{\kappa}{X}(X-1)\left(1-\frac{X+1}{2Z}\right)=\frac{Z^2}{X^2},$$

so that the saliency update map in (12) is determined by

$$\begin{cases} G(Z,X)=\frac{1}{\kappa}, \\ Z^+=\frac{Z^2}{X^2}, \end{cases} \qquad X\in[1,Z), \tag{59}$$

where

$$G(Z,X):=-\frac{X(X-1)(X+1-2Z)}{2Z^3}. \tag{60}$$

For a given Z, the admissible value of X is the smallest solution of $G(Z,X)=1/\kappa$ in $[1,Z)$, if such a solution exists.

Existence of the iteration. We present a condition on Z ensuring that the iteration in (59) exists, namely that there exists a solution $X(Z)\in[1,Z)$ of the equation $G(Z,X)=1/\kappa$ (see also Fig. 7). Observe that $G(Z,1)=0$ and that $G(Z,Z)=\frac{1}{2}\left(\frac{Z-1}{Z}\right)^2$. This implies that, by taking Z such that

$$\frac{1}{\kappa}<\frac{1}{2}\left(\frac{Z-1}{Z}\right)^2 \quad\Longleftrightarrow\quad Z>\frac{1}{1-\sqrt{2/\kappa}} \tag{61}$$

a solution $X(Z)\in[1,Z)$ must exist. Such solution is unique. Indeed, since

$$\frac{\partial G}{\partial X}=-\frac{3X^2-4ZX-1+2Z}{2Z^3} \tag{62}$$

the sign of the this derivative coincides with the sign of the function $q_Z(X):=-(3X^2-4ZX-1+2Z)$ that is a concave parabola. Since $q_Z(1)=2(Z-1)$ and $q_Z(Z)=(Z-1)^2$ that are both positive, we have that $q_Z(X)$ is positive in the interval $[1,Z)$. Hence $G(Z,X)$ is strictly increasing in that interval, so whenever (61) holds there is a unique solution $X(Z)\in[1,Z)$ and $Z^+=Z^2/X(Z)^2$ is uniquely defined.

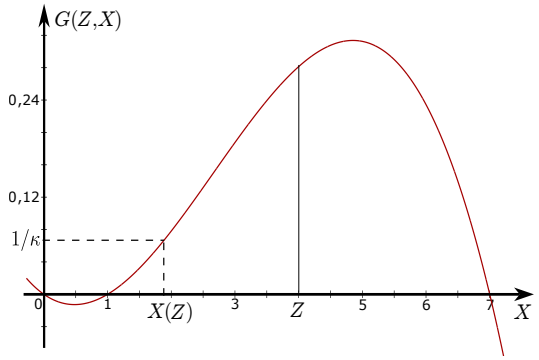

Figure 7: Plot of the function $G(Z,X)$ in (60) for $\kappa=15$ and $Z=4$.

Fixed points of the iteration. Fixed points satisfy $Z^+=Z$. Since $Z^+=Z^2/X^2$, this gives $Z=X^2$. Substituting into $G(Z,X)=1/\kappa$ yields (see also Fig. 8)

$$\frac{1}{\kappa}=G(X^2,X)=\frac{(X-1)^2(2X+1)}{2X^5} \tag{63}$$

Observe that

$$\frac{dG(X^2,X)}{dX}=-\frac{(X-1)}{2X^6}(4X^2-5X-5) \tag{64}$$

If we let $\bar{X} = \frac{5+\sqrt{105}}{8} \simeq 1.9059$, then we have that $G(X^2, X)$ is strictly increasing for $1 \leq X < \bar{X}$ the function, and strictly decreasing for $X > \bar{X}$, so that $G(X^2, X)$ reaches its maximum in $\bar{X}$. Since $G(\bar{X}^2, \bar{X}) \simeq 0.0785$, we can argue that we have fixed points only when

$$\kappa \geq \kappa_{\text{crit}} := \frac{1}{G(\bar{X}^2, \bar{X})} \simeq 12.74$$

For $\kappa > \kappa_{\text{crit}}$, there are exactly two solutions, $\bar{X}_1$ and $\bar{X}_2$, of (63) which satisfy $1 < \bar{X}_1 < \bar{X} < \bar{X}_2$. Such solutions in turn yield two fixed points, $\bar{Z}_1 = \bar{X}_1^2$ and $\bar{Z}_2 = \bar{X}_2^2$ which satisfy $1 < \bar{Z}_1 < \bar{Z} < \bar{Z}_2$ with $\bar{Z} = \bar{X}^2 = 3.6323$.

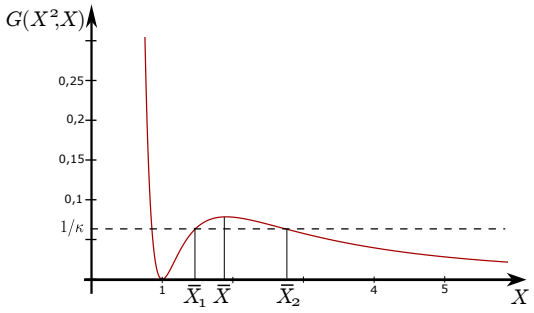

Figure 8: Plot of the function $G(X^2, X)$ in (63) and its intersections with the line $1/\kappa$.

Monotonicity of the iteration. We will now show that the iteration $Z^+ = \frac{Z^2}{X(Z)^2}$ is strictly increasing in Z, which proves that the sequence of iterates, denoted by $Z_0, Z_1, Z_2, \ldots$, is monotonic and hence it converges to a limit point Z_∞ . First notice that

$$\frac{\partial Z^+}{\partial Z} = \frac{2Z}{X^2} - \frac{2Z^2}{X^3}\frac{\partial X}{\partial Z} = \frac{2Z}{X^2} + \frac{2Z^2}{X^3}\frac{\partial G/\partial Z}{\partial G/\partial X} = \frac{2Z}{X^3}\frac{1}{\partial G/\partial X}\left(X\frac{\partial G}{\partial X} + Z\frac{\partial G}{\partial Z}\right)$$

Using

$$\frac{\partial G}{\partial Z} = \frac{X(X-1)(3X-4Z+3)}{2Z^4},$$

we obtain

$$X\frac{\partial G}{\partial X} + Z\frac{\partial G}{\partial Z} = \frac{X(Z-1)}{Z^3},$$

and hence

$$\frac{\partial Z^+}{\partial Z} = \frac{2(Z-1)}{X^2Z^2}\frac{1}{\partial G/\partial X}.$$

That shows that the sign of $\partial Z^+/\partial Z$ only depends on the sign of $\partial G/\partial X$. But, if the existence condition (61) holds, the derivative $\partial G/\partial X$ in (62) evaluated at $X(Z) \in [1, Z)$ must be positive. Thus, we conclude that iteration $Z^+ = \frac{Z^2}{X(Z)^2}$ is strictly increasing if $Z > 1/(1 - \sqrt{2/\kappa})$.

Convergence of the iteration. We show finally that the fixed point $\bar{Z}_1$ is repelling (or unstable) and that $\bar{Z}_2$ is attractive (or stable). To this aim observe that, at the fixed points, where $Z = X^2$, we have

$$\frac{\partial Z^+}{\partial Z} = \frac{4(X+1)}{4X^2 - X - 1} \quad \Longrightarrow \quad \frac{\partial Z^+}{\partial Z} - 1 = -\frac{4X^2 - 5X - 5}{4X^2 - X - 1}.$$

Then $\frac{\partial Z^+}{\partial Z} > 0$. Moreover, comparing the expression of $\frac{\partial Z^+}{\partial Z} - 1$ with (64), we see that, since $4X^2 - X - 1 > 0$, then $\frac{\partial Z^+}{\partial Z} - 1$ has the same sign of $\frac{dG(X^2,X)}{dX}$ showing in this way that $\frac{\partial Z^+}{\partial Z} > 1$ in $X = \bar{X}_1$ and $0 < \frac{\partial Z^+}{\partial Z} < 1$ in $X = \bar{X}_2$. This proves that the fixed point $\bar{Z}_1$ is repelling while the fixed point $\bar{Z}_2$ is attractive.

The previous analysis, together with the monotonicity of the iteration, shows that amplitudes above the unstable threshold $\bar{Z}_1$ converge to the attracting fixed point $\bar{Z}_2$, which corresponds to self-sustained sequential retrieval. In contrast, amplitudes below $\bar{Z}_1$ decay until subsequent transitions can no longer be triggered, ultimately leading the dynamics to collapse into the inactive state. This is illustrated in Fig. 4 in the main text.

E Additional results for Tanh activation

We complement the analysis of Sec. 4 by considering the Tanh activation $\psi(x) = \tanh(x)$, while retaining $\phi(z) = |z|$. Unlike SmoothF, the corresponding saliency map and escape-time predictions do not admit simple closed-form expressions, but can be computed numerically from the implicit equations derived in Sec. 4. Fig. 9 illustrates the resulting discrete map and representative saliency dynamics for $\kappa = 15$.

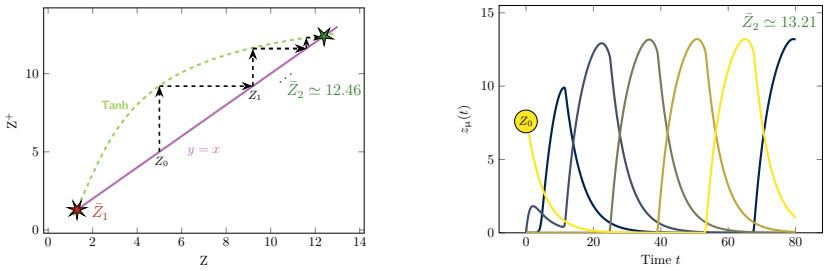

Figure 9: Discrete iteration map $Z^+ = h(Z/X)$ for the $\psi = \tanh$ activation function and associated z-components for the case of $\kappa = 15$.

We also repeat the numerical experiments of Fig. 5 using Tanh activation (Fig. 10). For orthogonal memories, theoretical and numerical escape times remain in good agreement (A, B), while increasing the noise amplitude reduces the switching delay and eventually leads to premature switching (C). For non-orthogonal memories, retrieval reliability decreases with memory bias and exhibits a non-monotonic dependence on the gain κ, with excessively large gains reducing the probability of successful transitions (D). Conditional on successful transitions, both the asymptotic escape time and saliency generally increase with κ (E, F). These results support the applicability of the theoretical framework to different activation functions, while highlighting activation-dependent differences in retrieval robustness.

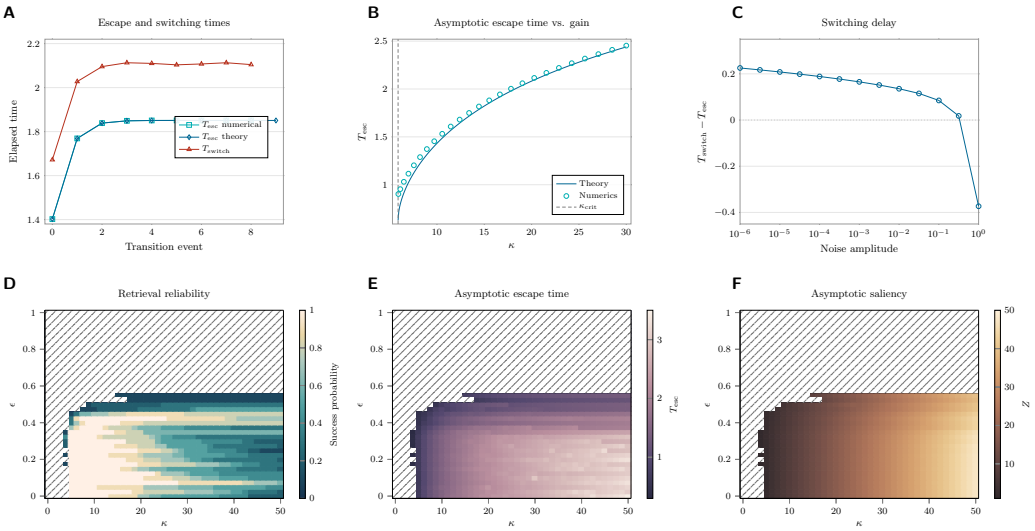

Figure 10: Numerical validation and robustness analysis for Tanh activation. (A) Theoretical and numerical escape times and observed switching times across consecutive transitions. (B) Asymptotic escape time versus κ: theory and numerical estimates. (C) Mean switching delay versus Gaussian noise amplitude. (D) Probability of correct transitions in the (κ, ϵ) plane. (E, F) Asymptotic escape time and saliency amplitude, respectively. All experiments use the same numerical configuration as Fig. 5. Hatched regions in (D-F) indicate no successful transitions.

F Numerical investigation of sequential retrieval capacity

An important open question is how sequential retrieval reliability scales with the number of stored memories and the length of the retrieved sequence. In classical Hopfield networks, the synaptic matrix is fixed, whereas in the proposed slow-fast model it is dynamically modulated by the saliency variables:

$$W_{\text{Hopfield}} = \frac{1}{N}\sum_{\mu=1}^{P} \xi^{\mu}(\xi^{\mu})^{\top}, \qquad W(|z|) = \frac{1}{N}\sum_{\mu=1}^{P} |z_{\mu}|\, \xi^{\mu}(\xi^{\mu})^{\top}.$$

In the classical setting, all stored memories contribute simultaneously to cross-talk. In contrast, during sequential retrieval, only a small subset of saliency coefficients typically remains significant, while the others decay toward zero. Thus, the nominal memory load P/N need not reflect the number of memories effectively contributing to the synaptic dynamics at a given time.

To investigate this effect, we consider two numerical protocols and measure the probability of successfully completing an entire prescribed retrieval cycle, following the sequential retrieval criterion of [19, Eq. (4)]. Successful retrieval requires the prescribed sequence of memory transitions, rather than merely escaping from the current memory state. In the *growing-cycle* protocol, a directed cycle of length $L = \text{round}(P/4)$ is embedded among P stored random i.i.d. memories with $\mathbb{P}(\xi_i^{\mu} = \pm 1) = 1/2$, with $N \in \{256, 512\}$. In the *fixed-cycle* protocol, the cycle length is kept at $L = 5$ while the memory load P/N is increased substantially, with $N = 256$. Both protocols are evaluated with SmoothF and Tanh activations ($\kappa = 15$ and $Z_0 = 8$), using 10 independent memory realizations per configuration.

Figure 11 reports the resulting retrieval probabilities. In the growing-cycle setting (panels A,B), SmoothF achieves reliable retrieval $P/N \in (17.5, 25)$ and $N = 512$, while retrieval probability constantly decreases for $N = 256$ over the studied memory load range. Tanh exhibits greater variability, particularly for smaller network sizes. In contrast, with a fixed cycle (panels C,D), retrieval remains highly reliable even for $P/N \gg 1$, with occasional failures for Tanh. This difference reflects, in part, the increasing number of correct transitions required when L grows with P, whereas the additional memories in the fixed-cycle setting remain largely inactive under the saliency dynamics.

These results highlight the distinction between nominal memory load and the number of memories actively involved in sequential retrieval. In particular, reliable retrieval of a fixed cycle at high P/N does not imply that all stored memories can be reliably retrieved. Rather, it suggests that dynamic saliency modulation can mitigate interference from memories outside the active sequence. Moreover, since both memory load and sequence length increase in the growing-cycle protocol, their individual effects cannot be isolated from these experiments. A precise theoretical characterization of sequential retrieval capacity therefore remains an open problem.

G Branching memory transitions

Our theoretical analysis focuses on transition graphs with at most one outgoing edge per memory, which enables a scalar reduction of the slow saliency dynamics and explicit conditions for sequential retrieval. However, this restriction is not inherent to the proposed model, which also accommodates branching graphs with multiple possible successors. In this setting, several successor saliencies are driven simultaneously, and branch selection depends on their competing dynamics.

To investigate this behavior, we consider two directed three-memory cycles sharing a common source memory, with equal outgoing transition weights. We use the SmoothF activation, set $N = 256$, $P = 5$, $\kappa = 15$, and generate 50 independent sets of random i.i.d. ± 1 memories. In the deterministic setting, the selected successor is the one with the larger absolute overlap with the source memory in all 45 realizations with unequal overlaps. We then introduce noise of amplitude 10^{-1} per integration step in the slow dynamics and perform 20 trials per memory realization. The successor with the larger absolute overlap remains the more frequently selected branch in 36 out of 43 realizations with unequal overlaps.

These results indicate that branch selection can be influenced by the geometry of the stored memories, even when outgoing transition weights are identical. In particular, differences in memory overlaps can break the symmetry between competing successors and favor one transition over another. This preference persists in most noisy trials, although stochastic perturbations can alter the selected branch.

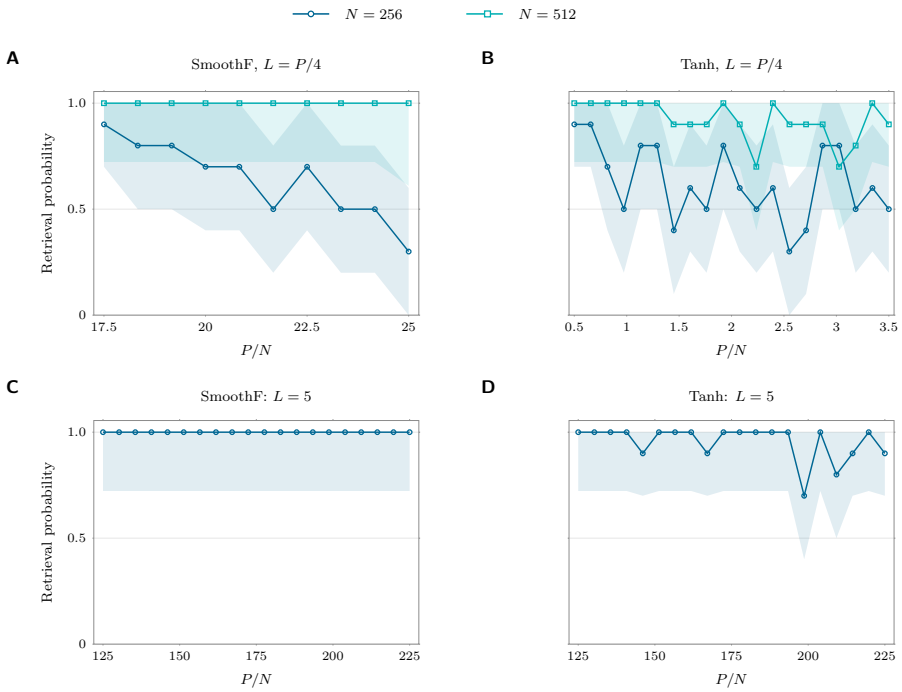

Figure 11: Full-sequence retrieval probability versus memory load P/N. (A, B) Growing cycles with $L \simeq P/4$ and $N \in \{256, 512\}$ for SmoothF and for Tanh, respectively. (C, D) Fixed cycles with $L = 5$ and $N = 256$, for SmoothF and Tanh, respectively. Shaded regions indicate 95% confidence intervals across independent memory realizations.

Thus, branching introduces a competition between successor memories that depends not only on the transition graph, but also on memory correlations and perturbations. A general characterization of this competition would require extending the scalar slow reduction to account for the simultaneous evolution of multiple successor saliencies.

H A motivating example for input-controlled transitions: Associative Memory Finite State Machine (AM-FSM)

Finite state machines (FSMs) are canonical models of sequential computation, in which the evolution of the system is jointly determined by current inputs and a compact internal representation of past states. At any time, an FSM is in one state drawn from a finite set, and incoming inputs trigger transitions according to a predefined state-dependent transition structure. In this sense, FSMs implement a form of implicit memory: past information is compressed into discrete internal states that condition future behavior. The state-dependent interpretation of inputs makes FSMs natural abstractions for a wide range of engineered systems, including elevators, network protocols, and traffic lights with pedestrian signals. In all such cases, effective control requires integrating current observations with the system's internal configuration, thereby generating context-appropriate actions over time. More broadly, decision-making processes exhibit a similar structure: actions depend not only on present stimuli, but also on the internal state shaped by prior events. This observation motivates viewing sequential decision-making as the evolution of an FSM.

As an illustrative example, consider a train driver operating a metro line in the Tokyo subway system. Under normal conditions, the driver repeatedly traverses a fixed route (the "red-black" line), corresponding to a stable sequence of internal states. Due to an unexpected strike, however, the driver receives instructions to deviate from the habitual route and instead pass through Shinjuku station (on the "green-black" line) and subsequently Tokyo station (on the "blue-black" line). Only two transfer hubs permit switching from the red-black line to the other lines. At each hub, the driver must choose

between alternative turns, using prior knowledge of the network and the current location to guide the decision. Although the map is outdated and uncertainty remains, the driver assigns high confidence to a left turn at each hub, while maintaining a lower probability for the alternative right turn.

This scenario exemplifies sequential decision-making under uncertainty, where each choice depends on both external cues and an internally maintained representation of the route taken so far. In the figure below, we instantiate the driver's behavior using a two-timescale IDP Hopfield model, illustrating how stable internal states and slower adaptation processes jointly support FSM-like dynamics. In particular, we instantiate an input-driven temporal network over the transition graph, where each outgoing edge from a memory ξ^ν towards a set of memories $\{\xi^\sigma\}_\sigma$ is suitably weighted by an external input u. The magnitude of the input components along each transition edge will ultimately determine the sequential memory path that is traversed by the two-timescale IDP Hopfield model. The exemplified control over sequential memory transitions provides solid insp

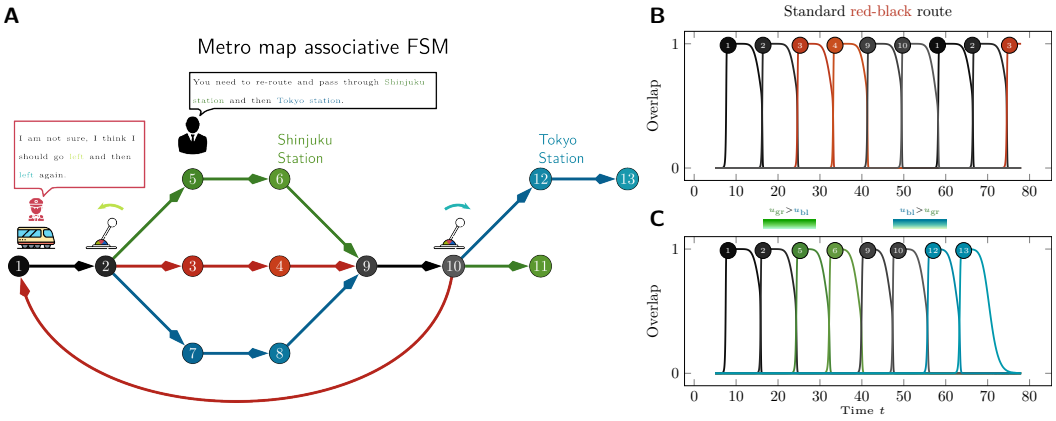

Figure 12: *Finite state machine dynamics with uncertain inputs realized by a two-timescale IDP Hopfield model.* (A) Schematic of a simplified Tokyo subway graph comprising three metro lines connected by two hub stations. The metro driver (red marker) must adapt the route in response to an external rerouting request, while facing uncertainty at each hub regarding the correct branch to take. (B) Sequential retrieval of stable memory states corresponding to stations along the habitual red-black line, illustrating deterministic FSM dynamics under standard operating conditions. (C) Input- and state-dependent sequential retrieval during rerouting: external inputs bias the transition dynamics at each hub toward the green-black line (via Shinjuku station) and subsequently the blue-black line (via Tokyo station), while preserving residual uncertainty across alternative transitions.

I Experimental setup

The code of the experimental campaign is available at `https://github.com/sim1bet/Memory-Flows-NeurIPS2026`.

We integrate the fast neural state and slow saliency dynamics by explicit Euler steps on the dimensionless fast clock. For $\tau_x/\tau_z = 0.01$, an integration step Δs corresponds to slow time $\Delta\theta = 0.01\Delta s$; the numerical step and horizon are specified separately for each experiment in `configs/paper/` and fast noise with amplitude $\sigma = 10^{-6}$.

The experiments are implemented in the accompanying repository through separate configuration-driven simulation and plotting scripts (`experiments/`, `plotting/`); the master command `run_all.sh` reproduces the configured workflows. For all experiments, successful numerical retrieval requires a persistent target alignment of at least 0.90, separated from the runner-up by at least 0.10, over a 0.05 slow-time hold window in the capacity tests. Uncertainty is computed over independent pattern sets, not over multiple source trajectories as if they were independent data sets.

All numerical campaigns for the camera-ready figures were executed on the *Leonardo* high-performance computing system, on the AI Factory project EHPC-AIF-2026PG02-1086. The repository includes resolved experimental configurations, seed policies, CSV outputs and figure-generation

scripts; numerical parallelism and worker counts are configurable through the command-line interface. Figure generation reads saved experimental data and does not rerun the underlying dynamics.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: All claims stated in the abstract and introduction are formally substantiated in the main paper. Specifically, Section 3 establishes the theoretical foundations of the proposed method, while Section 4 provides rigorous empirical and analytical validation of the main contributions. Complete proofs, detailed derivations, and auxiliary lemmas supporting these claims are deferred to Appendices B, C, and D to preserve readability of the main text. The scope of the contributions is further delimited in Section 5, where we explicitly discuss the assumptions underlying our framework, its current limitations, and directions for future work.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 5 of the main text provides a dedicated and structured discussion of the three principal limitations of the proposed framework. First, we identify the theoretical assumptions that would need to be relaxed in order to establish a tighter and more general parallelism with modern Transformer architectures, outlining concrete directions toward closing this gap. Second, we analyze the numerical sensitivity of the proposed simulation algorithm, characterizing the conditions under which excessive stability may arise and discussing potential mitigation strategies. Third, we examine the structural constraints imposed on the transition graph between memory states, clarifying the regimes in which these constraints are binding and their implications for the generality of our results.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.

- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: Every theoretical result presented in the main text is stated with its full set of assumptions and accompanied by a complete proof. To preserve the flow of the main text, detailed proofs are deferred to the Appendix, where each result is restated in full before its proof for ease of reference. Beyond the proofs of the main results, the Appendix also develops a set of supplementary lemmas and auxiliary results that serve to make the theoretical development self-contained, and to provide the reader with the technical background necessary to follow the core derivations.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper provides full disclosure of all information required to reproduce the experimental results reported in the main text. Specifically, Section 4 and Appendix F details the experimental setup, including all relevant parameters, initial conditions, and simulation procedures. The theoretical predictions against which the experimental results are validated are derived in full in Section 4 of the main text and Appendix D, where all intermediate steps are made explicit. Appendix D further consolidates the complete set of closed-form expressions and algorithmic specifications needed to independently simulate the system and verify the theoretical predictions. Together, these sections provide a self-contained and reproducible account of all results on which the main claims and conclusions of the paper rest.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.

- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The simulation code necessary to reproduce all experimental results are included in the supplemental material, alongside the parameter specifications detailed in Appendix D. The Ti*k*Z code used for figure rendering is not included, as it does not affect the reproducibility of the underlying results.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: The experimental framework of the paper does not involve a learned model and therefore does not entail training in the conventional sense: there is no optimizer, loss function, or data split. The experimental setting consists of a numerical simulation of coupled slow-fast dynamical systems, for which reproducibility is governed by the full specification of the system parameters rather than training hyperparameters. All such parameters - including the coupling coefficients, timescale separation constants, gain conditions, integration scheme, step size, and simulation horizon - are disclosed in full in Section 4 of the main text and Appendix D, ensuring that the simulations can be independently reproduced and the theoretical predictions verified.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: The paper is primarily theoretical in nature and its main contributions consist of analytical results derived from first principles. The experimental section validates these analytical predictions through deterministic numerical simulations of the proposed dynamical system, whose outcomes are fully determined by the system parameters and initial conditions rather than by stochastic processes. As such, the notions of error bars, confidence intervals, and statistical significance tests are not applicable to the results reported in this paper. The agreement between theoretical predictions and simulation outcomes is assessed directly through analytical comparison rather than statistical testing, which is the appropriate validation methodology for this class of results.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: Full details of the computational resources required to reproduce all experiments are provided in Appendix F. Specifically, all simulations were run on a Lenovo ThinkPad X1 Carbon equipped with an Intel Evo CPU and 32 GB of RAM, with no GPU acceleration. Given the deterministic nature of the forward Euler integration scheme and the modest scale of the simulations ($N = 256$ units, $P = 15$ memories, $T = 8000$ steps with $\Delta t = 0.01$), all experiments are well within the capacity of a standard laptop CPU and require no specialized hardware or distributed computing infrastructure to reproduce.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: The research presented in this paper is purely theoretical and computational in nature, consisting of mathematical analysis and numerical simulation of dynamical systems for sequential memory retrieval. No human subjects, personal data, sensitive datasets, or real-world deployment systems are involved at any stage of the research. The work poses no foreseeable risks of harm, dual use, or misuse, and raises no issues related to fairness, privacy, or discrimination. The authors have reviewed the NeurIPS Code of Ethics in full and confirm that the research conforms with its guidelines in every respect.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: The primary positive societal impact of this work lies in advancing the mechanistic understanding of sequential information flow in memory retrieval systems. By providing a rigorous theoretical framework for this phenomenon, the paper contributes foundational insights that may inform the interpretation and analysis of information flow in modern large language models, where analogous sequential processing mechanisms are poorly understood at a mechanistic level. Beyond this scientific contribution, the work is purely theoretical in nature and carries no foreseeable negative societal impacts: it involves no deployment of AI systems, no use of sensitive or personal data, and no application domain in which misuse or harm could plausibly arise.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.

- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This question is not applicable to the present work. The paper releases no pretrained models, datasets, or any other artifacts that carry a risk of misuse. The sole research outputs are mathematical derivations, theoretical results, and the numerical simulations of a dynamical system, none of which require safeguards for responsible release.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All external assets used in this paper are properly credited. No proprietary software, private datasets, or third-party models are used at any stage of the research. All licenses are permissive and fully compatible with academic publication.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.

- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: This question is not applicable to the present work. The paper introduces no new assets: no datasets, pretrained models, or software packages are released alongside the paper. The only research outputs are theoretical results and numerical simulations, which are fully documented within the paper and its Appendix.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This question is not applicable to the present work. The paper involves no crowdsourcing experiments or research with human subjects of any kind. All results are derived analytically or obtained through numerical simulation of a dynamical system.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

Justification: This question is not applicable to the present work. The paper involves no human subjects, study participants, or data collected from individuals. As a purely theoretical and computational study, no risk disclosure, informed consent procedure, or IRB approval was required or sought.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: This question is not applicable to the present work. No LLMs were used as a component of the core methodology, scientific derivations, or numerical simulations. LLM assistance was limited to writing support and the implementation of standard numerical utilities, neither of which contributes to the originality or scientific rigor of the research.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.