A DYNAMICAL THEORY OF SEQUENTIAL RETRIEVAL IN INPUT-DRIVEN HOPFIELD NETWORKS

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

Reasoning is the ability to integrate internal states and external inputs in a meaningful and semantically consistent flow. Contemporary machine learning (ML) systems increasingly rely on such sequential reasoning, from language understanding to multi-modal generation, often operating over dictionaries of prototypical patterns reminiscent of associative memory models. Understanding retrieval and sequentiality in associative memory models provides a powerful bridge to gain insight into ML reasoning. While the static retrieval properties of associative memory models are well understood, the theoretical foundations of sequential retrieval and multi-memory integration remain limited, with existing studies largely relying on numerical evidence. This work develops a dynamical theory of sequential reasoning in Hopfield networks. We consider the recently proposed input-driven plasticity (IDP) Hopfield network and analyze a two-timescale architecture coupling fast associative retrieval with slow reasoning dynamics. We derive explicit conditions for self-sustained memory transitions, including gain thresholds, escape times, and collapse regimes. Together, these results provide a principled mathematical account of sequentiality in associative memory models, bridging classical Hopfield dynamics and modern reasoning architectures.

1 INTRODUCTION

Associative memory has been a foundational concept in theoretical neuroscience and machine learning for over four decades. In its classical formulation, memory retrieval is modeled as convergence toward stable equilibria of an energy landscape (Hopfield, 1982; 1984), 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 (Amit et al., 1987; McEliece et al., 1987; Tsodyks & Feigel'man, 1988; Betteti et al., 2024), revealing a fundamental limitation of the original Hopfield network, whose capacity scales sub-linearly with the number of neurons. Modern Hopfield models (Krotov & Hopfield, 2016; 2020) substantially extend this paradigm, achieving exponential memory capacity while retaining an energy-based formulation (Demircigil et al., 2017). Beyond capacity, these models admit strong interpretative links to contemporary machine learning architectures, including Transformers (Ramsauer et al., 2021; Hoover et al., 2023) and generative diffusion models (Ambrogioni, 2024). Despite these advances, modern Hopfield networks largely remain static retrieval systems: dynamics halt upon convergence to a memory, and sequential retrieval requires repeated reinitialization. This limitation contrasts with many cognitive and algorithmic tasks, including those addressed by Transformers, which rely on structured transitions across sequences of memories, and motivates the need for dynamical mechanisms that induce controlled transitions between energy minima to support reasoning over time.

Early work by Kleinfeld (1986) already identified this limitation and proposed a continuous-time mechanism for sequential memory retrieval via delayed asymmetric interactions that drive transitions between attractors. This idea has since motivated extensive efforts to model memory transitions, with contemporary work (Sompolinsky & Kanter, 1986) establishing mean-field results on the gain sustaining sequentiality. Recently, several works have revisited sequential retrieval within the modern Hopfield formalism (Karuvally et al., 2023; Chaudhry et al., 2024; Karuvally et al., 2025), typically by introducing slow variables or delayed neuronal layers that modulate synaptic couplings over

time. Related approaches employ slowly evolving parameters to control attractor switching (Herron et al., 2023). While effective in simulations, these augmented dynamics significantly reduce analytical tractability, obscuring the mechanistic understanding that made the original Hopfield model appealing.

In this paper, we build on recent results on input-controlled transitions in Hopfield networks (Betteti et al., 2025) to provide a complete dynamical characterization of sequential reasoning in modern Hopfield models. We introduce an input-driven plasticity (IDP) Hopfield framework that exploits a separation of timescales across architectural layers to induce predictable and analytically tractable transitions between memories. This structure enables derivation of exact conditions for memory escape times and gain parameters required to sustain sequential flows. By unifying dynamics and geometry, our framework provides a mechanistic account of how modern Hopfield architectures can implement sequential memory retrieval and support structured reasoning. Figure 1 provides a schematic overview of the sequential memory transition mechanism introduced in this work.

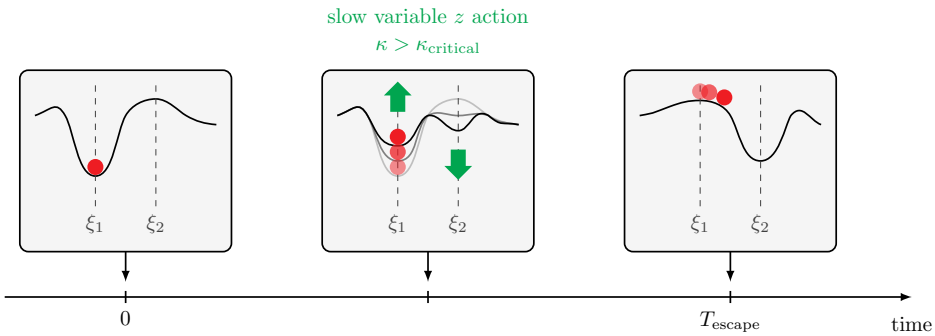

Figure 1: **Schematic illustration of the proposed two-timescale mechanism for sequential memory transitions.** At the initial time, the fast state variable is at a stable equilibrium corresponding to memory ξ_1. The slow reasoning variable z progressively destabilizes this equilibrium while simultaneously stabilizing the equilibrium associated with memory ξ_2. At the escape time T_{escape}, 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_{critical}. We analytically characterize both k_{critical} and the resulting escape time T_{escape}.

2 Retrieval: the IDP Hopfield model and its modern Hopfield formulation

Classical associative memory models describe retrieval as a fast relaxation process, whereby a slowly varying external or cortical input cues reconstruction of a stored pattern. In most formulations, however, the explicit dependence of retrieval dynamics on the input is neglected, leading either to sensitivity to initial conditions or to distortions of the stored memories, and limiting algorithmic interpretability. To address this limitation, recent work introduced the input-driven plasticity (IDP) Hopfield model (Betteti et al., 2025), an extension of the classical Hopfield network in which synaptic interactions depend multiplicatively on a filtered input. This formulation enables a precise characterization of the input–memory relationship, including equilibrium existence and input-dependent stability.

The IDP Hopfield dynamics are given by

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

where $\alpha = \alpha(u) \in \mathbb{R}^P$ encodes a filtering of a slow-varying input u and $W(\alpha) \in \mathbb{R}^{N\times N}$ is a synaptic matrix constructed from stored memories $\xi^1, \dots, \xi^P \subseteq \mathbb{R}^N$. The classical Hopfield model is recovered when all $\{\alpha_\mu\}_{\mu=1}^P$ are equal and positive. Details of the construction are deferred to the Appendix.

When the activation Ψ is the gradient of a convex function (Krotov & Hopfield, 2020), the dynamics admit a quasi-gradient representation

$$\dot{x} = -\frac{\partial \Psi}{\partial x}(x)^{-1}\nabla \mathrm{E}(x, \alpha), \tag{2}$$

and in particular, when Ψ is diagonal, the associated energy becomes

$$\mathrm{E}(x, \alpha) = -\tfrac{1}{2}\Psi(x)^\top W(\alpha)\Psi(x) + x^\top \Psi(x) - \sum_{i=1}^{N} \int_0^{x_i} \psi(s), ds. \tag{3}$$

The energy guarantees convergence to stable equilibria and provides a geometric interpretation of memories and their basins of attraction. Each coefficient α_ν controls the existence and stability of the corresponding memory: the ν-th memory exists as an equilibrium if and only if $\alpha_\nu > \alpha_{\text{existence}} = 1$, and is stable when $\alpha_\nu > \alpha_{\text{stability}}$ exceeds a stability threshold determined by the largest entry of α.

While the IDP Hopfield model captures input-controlled transitions between memories, it does not address the structure of transitions or semantic relations among memories. To this end, we consider a modern Hopfield generalization that introduces a separation of timescales across three interacting layers:

$$\tau_y \dot{y} = -y + M_y \Psi_y(x) \qquad \boxed{\text{memory layer}}, \tag{4}$$

$$\tau_x \dot{x} = -x + M_x[z \odot \Psi_x(z \odot y)] \qquad \boxed{\text{feature layer}}, \tag{5}$$

$$\tau_z \dot{z} = -z + M_z \Psi_z(u) \qquad \boxed{\text{saliency layer}}, \tag{6}$$

Under the timescale separation $\tau_z, \tau_x \gg \tau_y$, the memory layer rapidly indexes stored patterns, the feature layer integrates internal memory structure with reasoning signals, and the slow reasoning layer accumulates evidence from external inputs. This architecture supports continuous, input-modulated memory retrieval and provides the basis for the sequential reasoning dynamics analyzed in the remainder of the paper. Specifically, in the remainder of the paper we will investigate the joint role of feedback modulation $u = x$ and of a sequential transition matrix as first presented in Kleinfeld (1986), and the saliency layer will henceforth be referred to as *reasoning layer*.

3 Reasoning: the two-timescale IDP Hopfield model

Most implementations of modern Hopfield models exploit a timescale separation between memory and feature dynamics to recover a single-timescale system that instantiates different retrieval mechanisms depending on the choice of activation function, e.g. Transformer-like retrieval for $\Psi_x(x) = \mathrm{softmax}(x)$ (Ramsauer et al., 2021). Under this limit, the model reduces to a standard associative memory system. In our tripartite architecture, we retain in addition a slow reasoning variable to construct a minimal two-timescale Hopfield model capable of autonomous sequential retrieval. We specialize our model (4), (5), (6) to the choices $\tau_y \to 0$, $M_y^\top = M_z^\top = M_x = M = N^{-1/2}[\xi^1, \dots, \xi^P]$, $\Psi_x = \mathrm{Id}$ and $\Psi_y = \Psi_z = \Psi$. Then the reasoning Hopfield model is defined by the coupled dynamics

$$\begin{cases} \tau_x \dot{x} = -x + M \mathrm{diag}(\alpha) M^\top \Psi(x), & x(0) \in \mathbb{R}^N, \\ \tau_z \dot{z} = -z + \frac{\kappa}{\sqrt{N}} A M^\top \Psi(x), & z(0) \in \mathbb{R}^P, \ \alpha = z \odot z, \end{cases} \tag{7}$$

where $\kappa > 0$ is a gain parameter and $\Psi(x) = (\psi(x_1), \cdots, \psi(x_N))$ is a diagonal, non-decreasing and saturating activation function. In the following, we consider the HardTanh activation function $\psi(z) = \max\{-1, \min\{x, 1\}\}$. The sequential associations between memories are instead coded by $A \in \mathbb{R}^{P \times P}$, which we name *reasoning matrix*. In our case A is a circulant matrix, i.e., $A_{\nu\mu} = \delta_{\nu+1,\nu}$, where $\delta_{i,j}$ denotes the Kronecker delta, and induces a limit cycle over the memories (see Fig. 2A). The circulant matrix A plays a role analogous to the transition matrix

$$Q = \frac{1}{N} \sum_{\mu=1}^{P-1} \xi^{\mu+1} \xi^{\mu\top} + \frac{1}{N} \xi^1 \xi^{P\top} \in \mathbb{R}^{N \times N} \tag{8}$$

$$= M A M^\top \tag{9}$$

introduced in the seminal work of Kleinfeld (1986), while operating at the level of memory indices rather than feature space. Comparing the dynamics in (7) with the Kleinfeld-type, one timescale Hopfield model $\dot{x} = -x + \kappa Q \Psi(x)$, we observe important differences in the sequential retrieval. As shown in Fig. 2B-C, the one-timescale model induces oscillatory behavior across memory components, but exhibits mixed states with low overlap, irregular escape times, and strong sensitivity to the gain κ. Depending on κ, the dynamics may collapse to the origin, converge to a fixed point, or sustain a limit cycle, consistent with previous observations (Gillett & Brunel, 2024). In contrast, in Fig. 2D-E the two-timescale reasoning model displays robust and predictable sequential retrieval for all $\kappa > 4$. Transitions are sharp, overlaps reach their maximal value, and escape times are uniform across memories, yielding clean transitions with exact periodicity.

3.1 Slow reasoning dynamics: the driving engine

The reasoning dynamics z evolve on the slowest timescale τ_z and fully govern memory transitions through the adjacency matrix $A \in \mathbb{R}^{P \times P}$. While previous two-timescale Hopfield models typically invoke transitions for sufficiently

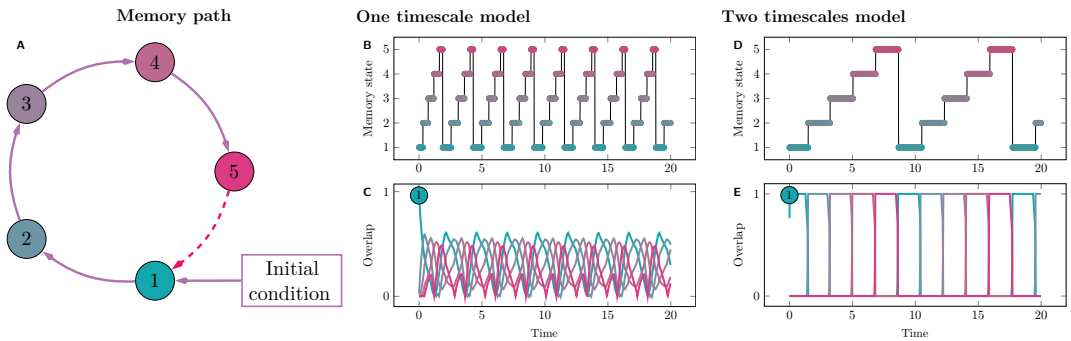

Figure 2: **Cyclic memory transitions in one- and two-timescale Hopfield models.** (A) Limit-cycle structure induced by the transition matrices Q in the one-timescale dynamics and A in the two-timescale reasoning dynamics (7), with both systems initialized near memory ξ^1. (B-C) One-timescale model: sequential retrieval proceeds through mixed memory states, with partial overlaps and unpredictable escape times across memories. (D-E) Two-timescale reasoning model: transitions occur between individual memories without mixing, with uniform, quantifiable escape times and overlaps reaching their maximal value, indicating exact memory alignment.

large gains Herron et al. (2023), the joint role of the gain parameter κ and the initial saliency configuration $z(0)$ has not been systematically characterized. Here we provide an explicit, coefficient-wise description of the slow dynamics and derive the exact condition determining whether transitions occur or the system collapses to inactivity. In what follows, we consider a HardTanh activation function and a circulant reasoning matrix A. We present the main results in concise form, with extended derivations deferred to Appendix A.2.

Assume that the fast feature dynamics rapidly converge to a memory equilibrium, that is

$$x_\nu^\star = z_\nu^2\, \xi^\nu \quad \text{with} \quad z_\nu^2 > 1. \tag{10}$$

Exploiting the fast convergence of the feature layer, the slow dynamics reduce to

$$\dot{z} = -z + \frac{\kappa}{\sqrt{N}} A M^\top \Psi(x_\nu^\star), \tag{11}$$

and using (10) and the fact that $\Psi(x_\nu^\star) = \xi^\nu$, equation 11 can be expressed component-wise as

$$\begin{cases} \dot{z}_{\nu+1} = -z_{\nu+1} + \kappa, \\ \dot{z}_\mu = -z_\mu \qquad \forall \mu \neq \nu + 1. \end{cases} \tag{12}$$

Thus, all saliency weights except $z_{\nu+1}$ decay exponentially, while $z_{\nu+1}$ is driven by a constant input κ. More precisely, assume that at the initial time $t = 0$, $z_\nu(0) = Z_0 > 1$, while all other $z_\mu(0)$, $\mu \neq \nu$ are negligible and, for simplicity, set to zero. Then, from (12),

$$z_\nu(t) = Z_0 e^{-t}, \quad z_{\nu+1}(t) = \kappa \int_0^t e^{-(t-\tau)} \mathrm{d}\tau = \kappa(1 - e^{-t}). \tag{13}$$

A transition from memory ξ^ν to $\xi^{\nu+1}$ occurs if two conditions are met: (i) the current equilibrium in the fast feature dynamics $x_\nu^\star$ loses stability, and (ii) a stable equilibrium $x_{\nu+1}^\star$ associated with the next memory $\xi^{\nu+1}$ appears. In energetic terms, condition (ii) requires the formation of a sufficiently deep well for the memory $x_{\nu+1}^\star$ where the fast dynamics can roll down to, compatibly with the example in Fig. 1.

From the stability analysis of the IDP Hopfield model, the equilibrium $x_\nu^\star$ loses stability at the time T^0_{escape} such that $Z_0 e^{T^0_{\text{escape}}} = 1$, that is $T^0_{\text{escape}} = \log Z_0$. In the specific case of the HardTanh, the loss of stability for the memories of the IDP Hopfield models happens as they leave the saturation regions of the activation function, hence for $z_\nu^2 \to 1$. Evaluating $z_{\nu+1}(t)$ in (13) at time T^0_{escape} provides the initial condition Z_1 for the dominant saliency weight in the next transition

$$Z_1 = z_{\nu+1}(T^0_{\text{escape}}) = \kappa(1 - 1/Z_0).$$

From the properties of the IDP Hopfield model in the case of the HardTanh activation function, the next memory equilibrium $x_{\nu+1}^\star$ exists and is stable if $Z_1 > 1$; otherwise, the activity collapses to the origin. Iterating the above argument over successive transitions leads to the discrete-time map

$$Z_{t+1} = \kappa \underbrace{(1 - 1/Z_t)}_{\mathcal{P}(Z_t)}. \tag{14}$$

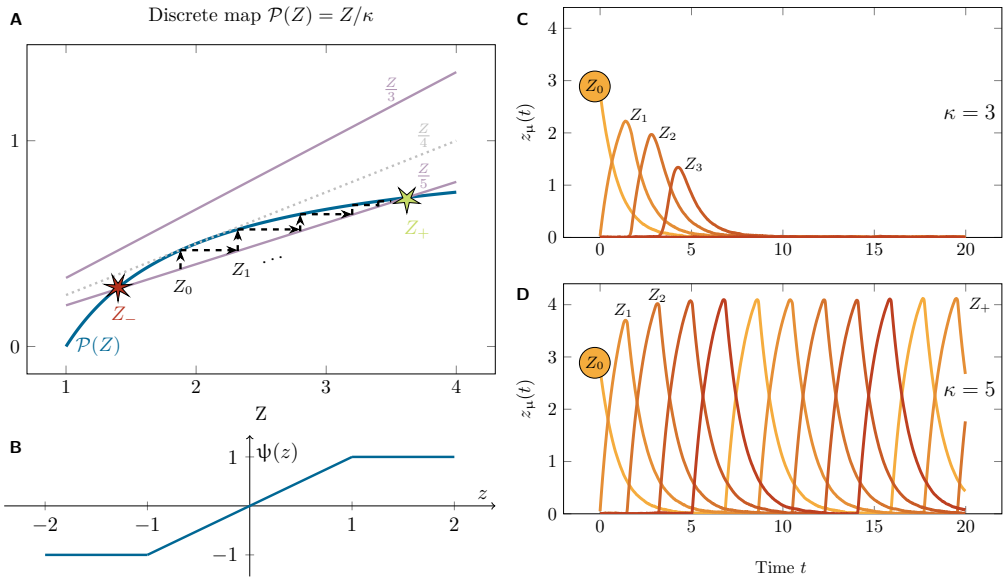

Figure 3: **Dependence of the slow reasoning dynamics on the gain κ and the initial condition Z_0.** (A) Solutions of the discrete iteration $\mathcal{P}(Z) = Z/\kappa$ as a function of the gain κ. Real solutions $Z_\pm > 1$ exist only for $\kappa \geq \kappa_{\text{critical}} = 4$ and sequential transitions are self-sustaining only if the initial saliency weight satisfies $Z_0 > Z_-$. (B) HardTanh activation function, linear in $(-1, 1)$ and saturated outside this interval. (C-D) Component-wise slow reasoning dynamics for different gain regimes. (C) Subcritical gain $\kappa = 3$: oscillations decay and the dynamics collapse to the origin, leading to loss of retrieval. (D) Supercritical gain $\kappa = 5$: oscillations persist and entrain to the stable fixed point Z_+, enabling reliable sequential memory retrieval.

which governs the peak values of the dominant saliency weights. Periodic, self-sustained transitions correspond to fixed points $Z > 1$ of this map. As shown in Fig. 3, such fixed points exist only for $\kappa \geq 4$, in which case they are given by

$$Z_\pm = \frac{\kappa \pm \sqrt{\kappa^2 - 4\kappa}}{2} > 1. \tag{15}$$

Moreover, for $\kappa \geq \kappa_{\text{critical}} = 4$, the trajectory $\{Z_t\}_{t\geq 0}$ converges monotonically to Z_+ if $Z_0 > Z_-$. We conclude that if $\kappa \geq 4$ and the initial condition satisfies $Z_0 > Z_-$[1], the system exhibits asymptotically periodic, self-sustained transitions with period $T^\infty_{\text{escape}} = \log Z_+$.

4 Conclusions

In this work, we developed a dynamical theory of sequential memory transitions in Hopfield-type models. By exploiting the analytical tractability of the IDP Hopfield framework, we introduced a two-layer, two-timescale retrieval–reasoning architecture that integrates fast associative recall with slow, structured transitions over learned associations. Sequentiality emerges entirely from the slow reasoning dynamics, yielding predictable memory escape times and explicit gain conditions required to sustain transitions.

In the setting of a HardTanh activation, we derived an exact gain threshold for self-sustained sequential retrieval. When this threshold is exceeded, the system exhibits stable, recurrent transitions between memories; below it, activity either collapses directly to the origin or undergoes transient, progressively weakening transitions before vanishing. These results provide a precise dynamical explanation of how sequential reasoning can arise in energy-based models without sacrificing analytical control.

Ongoing work extends this framework toward a geometric characterization of the memory manifolds supporting retrieval and of the fibers mediating transitions between them. By unifying geometric structure with timescale-separated dynamics, our aim is to establish a complete mathematical account of sequentiality in Hopfield-type networks, offering mechanistic insight into how modern energy-based architectures can support structured, multi-step reasoning.

[1] In the case $Z_0 < Z_-$ the discrete map $\mathcal{P}(Z) = Z/\kappa$ pushes $Z_\infty \to 1$, at which point the entire fast-slow system activity collapses to the origin.

REFERENCES

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.

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.

S. Betteti, G. Baggio, and S. Zampieri. On the capacity of continuous-time hopfield models. In *2024 IEEE 63rd Conference on Decision and Control (CDC)*, pp. 7853–7858. IEEE, December 2024. doi: 10.1109/cdc56724.2024.10886497.

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.

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.

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.

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.

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.

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.

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

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.

A APPENDIX

A.1 THE IDP HOPFIELD MODEL

The IDP Hopfield model presented in Betteti et al. (2025) is characterized by the O.D.E.

$$\dot{x} = -x + W(\alpha)\Psi(x) \tag{16}$$

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*

$$\Sigma := \{\xi^1, \dots, \xi^P\} \subseteq \{-1, 1\}^N \tag{17}$$

$$\xi^{\mu\top}\xi^{\nu} = N\delta_{\mu,\nu} \qquad \mu, \nu \in \{1, \dots, P\} \tag{18}$$

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

$$\mathrm{W}(\alpha) = \frac{1}{N}\sum_{\mu=1}^{P} \alpha_\mu \xi^\mu \xi^{\mu\top}. \tag{19}$$

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.

$$\mathbb{P}[\xi_i^\nu = 1] = \mathbb{P}[\xi_i^\nu = -1] = \frac{1}{2} \qquad \forall \nu = 1, \dots, P, \quad \forall i = 1, \dots, N \tag{20}$$

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

$$M = \frac{1}{\sqrt{N}}[\xi^1, \dots, \xi^P] \tag{21}$$

we can rewrite the input-modulated synaptic matrix as

$$W(\alpha) = M\mathrm{diag}(\alpha)M^\top \tag{22}$$

The canonical synaptic matrix W of the Hopfield model is a specific instance of equation 19 for the choice $\alpha = \mathbb{1}$ (or a scalar multiple of it).

We now specialize the treatment of the IDP Hopfield model as presented in Betteti et al. (2025) to the case of the HardTanh activation function. Specifically, we consider $\Psi : \mathbb{R}^N \to \mathbb{R}^N$ to be diagonal and homogeneous, i.e.,

$$\Psi_i(x) = \psi(x_i) \tag{23}$$

with $\psi(s) = \max\{-1, \min\{1, s\}\}$ for $s \in \mathbb{R}$.

A.1.1 EXISTENCE OF THE MEMORIES AS EQUILIBRIUM POINTS

Starting from the condition for the existence of the equilibria in Betteti et al. (2025), we gain an explicit parametrization of $x_\nu^\star = \gamma_\nu \xi^\nu$ when such equilibria exist. In particular, the saliency weights $\alpha_\nu = z_\nu^2$, for $\nu = 1, \dots, P$, must uniquely satisfy

$$\alpha_\nu = z_\nu^2 = \frac{\gamma_\nu}{\psi(\gamma_\nu)} \tag{24}$$

which in this specific case becomes

$$z_\nu^2 = \begin{cases} 1 & \gamma_\nu \in (0, 1) \\ \gamma_\nu & \gamma_\nu \in [1, +\infty) \end{cases} \tag{25}$$

If $z_\nu^2 < 1$, then there does not exist $\gamma_\nu > 0$ such that (24) is satisfied. Instead, for $z_\nu^2 = 1$ (24) is trivially satisfied by any $\gamma_\nu \in (0, 1]$. Finally, when $z_\nu^2 > 1$ it is easy to see that no $\gamma_\nu \in (0, 1]$ satisfies (24), and that for any $\gamma_\nu > 1$ we have

$$z_\nu^2 = \gamma_\nu \tag{26}$$

It then follows that existing equilibria are explicitly parametrized as $x_\nu^\star = z_\nu^2 \xi^\nu$.

A.1.2 Local stability of the memories as equilibrium points

The local stability condition for the memories requires a direct computation of the Jacobian of the Hopfield map $f(x, \alpha) = -x + W(\alpha)\Psi(x)$[2]. At the equilibrium points, we have

$$\begin{aligned} \frac{\partial f}{\partial x}(x^\star_\nu) &= -\boldsymbol{I} + W(\alpha)\frac{\partial \Psi}{\partial x}(x^\star_\nu) \\ &= -\boldsymbol{I} + W(\alpha)\psi'(z^2_\nu)\boldsymbol{I} \end{aligned} \tag{27}$$

where in the last passage we have exploited the symmetry of the derivative of the HardTanh with respect to the origin. In particular, we will have that $\psi'(z^2_\nu) = 0$ for any $z^2_\nu > 1$. Thus

$$\frac{\partial f}{\partial x}(x^\star_\nu) = -\boldsymbol{I} \qquad \forall z^2_\nu > 1 \tag{28}$$

and therefore the existence condition for the memories and their local stability condition coincide. In particular, we conclude that $x^\star_\nu$ exists and is a locally exponentially stable memory of the IDP Hopfield model with HardTanh activation function for all $z^2_\nu > 1$.

A.2 Slow reasoning dynamics: under the hood

Under the fast relaxation of the feature dynamics to the stable equilibrium $x^\star_\nu = z^2_\nu \xi^\nu$, we want to study the unfolding of the slow dynamics. Starting from the definition of the slow dynamics, and noticing that for HardTanh activation function $\Psi(x^\star_\nu) = \psi(z^2_2)\xi^\nu = \xi^\nu$, since $z^2_\nu > 1$, we get

$$\begin{aligned} \dot{z} &= -z + \frac{\kappa}{\sqrt{N}} A M^\top \Psi(x^\star_\nu) \\ &= -z + \frac{\kappa}{\sqrt{N}} A M^\top \xi^\nu \\ &= -z + \kappa A \boldsymbol{e}^{(\nu)} \\ &= -z + \kappa \boldsymbol{e}^{(\nu+1)}, \end{aligned} \tag{29}$$

where $\boldsymbol{e}^{(\ell)}$ denotes the ℓ-th vector of the canonical basis of $\mathbb{R}^P$. It then becomes clear that the study of the dynamics is reduced to the simple components condition

$$\begin{cases} \dot{z}_{\nu+1} = -z_{\nu+1} + \kappa, \\ \dot{z}_\mu = -z_\mu, \qquad \mu \neq \nu + 1 \end{cases} \tag{30}$$

and considering for simplicity initial conditions $z_\nu(0) = Z_0 > 1$, $z_\mu(0) = 0$ for all $\mu \neq \nu$ and integrating we get

$$\begin{cases} z_\nu(t) = Z_0 e^{-t} \\ z_{\nu+1}(t) = \kappa \int_0^t e^{-(t-\tau)}\, d\tau = \kappa(1 - e^{-t}) \end{cases} \tag{31}$$

In order for the transition to occur, we need the memory $x^\star_\nu$ to lose stability as given by (28). Therefore, there must exist a first escape time $T^0_{\text{escape}} > 0$ such that

$$1 = Z_0 e^{-T^0_{\text{escape}}} \tag{32}$$

which trivially gives $T^0_{\text{escape}} = \log Z_0$. Using the new estimate of the first escape time in (31) we obtain that

$$\begin{aligned} z_{\nu+1}(T^0_{\text{escape}}) &= \kappa(1 - e^{-\log Z_0}) \\ &= \kappa\left(1 - \frac{1}{Z_0}\right). \end{aligned} \tag{33}$$

Calling $Z_1 = z_{\nu+1}(T^0_{\text{escape}})$, we will have a new escape time for the memory $x^\star_{\nu+1}$ estimated as $T^1_{\text{escape}} = \log Z_1$, from which it will be possible to estimate $Z_2 = \kappa(1 - 1/Z_1)$ and so forth. This iterative mechanism yields the one-dimensional discrete-time dynamical system

$$\begin{aligned} Z_{t+1} &= \kappa\left(1 - \frac{1}{Z_t}\right) \\ &= \kappa \mathcal{P}(Z_t) =: \mathcal{F}(Z_t). \end{aligned} \tag{34}$$

[2]Notice that by the non-smoothness of the HardTanh activation function, the Jacobian is not defined for any $x \in \partial[-1, 1]^N$ boundary of the hypercube $[-1, 1]^N$. Since the existence of the memories requires $z^2_\nu > 1$, then we automatically have that $x^\star_\nu \notin \partial[-1, 1]$, and for the present treatment with do not consider singularities on measure zero sets.

Fixed points $Z > 1$ of (34) correspond to periodic saliency peaks of the components of z, i.e., to periodic self-sustained transitions. They are the solutions of $\mathcal{F}(Z) = Z$, equivalently the real roots of

$$Z^2 - \kappa Z + \kappa = 0. \tag{35}$$

Such roots are real and greater than 1 if and only if $\kappa \geq 4$, in which case

$$Z_\pm = \frac{\kappa \pm \sqrt{\kappa^2 - 4\kappa}}{2}. \tag{36}$$

To determine convergence, we examine the increment

$$Z_{t+1} - Z_t = \mathcal{F}(Z_t) - Z_t,$$

whose sign indicates whether the sequence increases or decreases. A direct computation gives

$$\mathcal{F}(Z) - Z = \kappa\left(1 - \frac{1}{Z}\right) - Z = -\frac{Z^2 - \kappa Z + \kappa}{Z} = -\frac{(Z - Z_-)(Z - Z_+)}{Z}.$$

For $Z > 1$, the denominator is positive, so the sign of $\mathcal{F}(Z) - Z$ is opposite to that of $(Z - Z_-)(Z - Z_+)$.

Case $\kappa > 4$. The fixed points are distinct and satisfy $1 < Z_- < Z_+$. Then

$$\mathcal{F}(Z) - Z \begin{cases} < 0, & 1 < Z < Z_-, \\ > 0, & Z_- < Z < Z_+, \\ < 0, & Z > Z_+. \end{cases}$$

Hence, if $1 < Z_0 < Z_-$ the sequence decreases toward the boundary $Z = 1$ and transitions are not sustained. If $Z_0 > Z_-$, the trajectory is monotone and remains trapped between Z_- and Z_+, converging to Z_+.

Case $\kappa = 4$. The two fixed points coincide at $Z_\pm = 2$, and

$$\mathcal{F}(Z) - Z = -\frac{(Z-2)^2}{Z} \leq 0.$$

Hence $Z_{t+1} \leq Z_t$ for all $Z > 1$, with equality only at $Z = 2$. If $Z_0 > 2$, the sequence decreases monotonically to 2. If $1 < Z_0 < 2$, it decreases away from 2 toward the boundary $Z = 1$, so sustained transitions do not occur.

In conclusion, for $\kappa \geq 4$ and $Z_0 > Z_-$, the model supports asymptotically periodic, self-sustained transitions, and the corresponding asymptotic period is given by

$$T^\infty_{\text{escape}} = \log Z_+. \tag{37}$$