# SEQUENTIAL PATTERN RETRIEVAL: NEW REPRESENTATIONS INSPIRED BY NON-EQUILIBRIUM PHYSICS AND ASSOCIATIVE MEMORY MODELS

**Matthew Du & Agnish Kumar Behera**
Department of Chemistry and The James Franck Institute
The University of Chicago
Chicago, IL 60637, USA
{madu,agnish}@uchicago.edu

**Madan Rao**
National Center for Biological Sciences
Bengaluru, India
madan@ncbs.ac.in

**Suriyanarayanan Vaikuntanathan**
Department of Chemistry and The James Franck Institute
The University of Chicago
Chicago, IL 60637, USA
svaikunt@uchicago.edu

## ABSTRACT

Generating a temporal sequence of outputs from a single output has broad relevance, including in neuroscience and machine learning. Inspired by ideas in non-equilibrium physics and modern associative memory models, we demonstrate new representations of sequence recall. Our findings provide potential strategies to improve the learning of temporal data in state-of-the-art neural networks.

## 1 INTRODUCTION

Machine learning models have been used productively to infer and model sequential data streams. Common examples of architectures used include recurrent neural networks (RNNs), such as long-short-term-memory (LSTM). Such neural networks have applications in speech recognition or motor control (Machens et al., 2005), where one phrase or action, respectively, leads to another. Associative memory-based models can provide an alternative (potentially unsupervised) framework for sequence generation and inference. The standard paradigmatic associative memory model, the so-called Hopfield model (Hopfield, 1982), is an energy-based model and can be viewed as a physical system with symmetric interactions. Starting from a corrupted pattern, the input is recurrently converted to a configuration of lower energy, ultimately reaching the energy minimum of the correct pattern. To enable dynamics other than energy descent to one pattern, in particular, the retrieval of multiple patterns in sequence, one can introduce asymmetric interactions (Hopfield, 1982; Yan et al., 2013). In particular, a simple mechanism for sequence retrieval in the Hopfield model was discussed in Kanter & Sompolinsky (1986), where asymmetric coupling is supplemented with a memory kernel in the dynamics. Inspired by modern Hopfield networks (Krotov & Hopfield, 2016), Chaudhry et al. (2024) employed nonlinear interactions to enable sequence retrieval without a memory kernel.

Inspired by nonequilibrium physics, we present two new representations of pattern cycling. First, using the well-known fact in nonequilibrium statistical mechanics that hidden degrees of freedom can be introduced to capture memory-like behavior, we show that the memory kernel of Kanter & Sompolinsky (1986) can be achieved using hidden neurons. Second, we demonstrate enhanced sequence retrieval in an asymmetric variant of the modern Hopfield networks by applying so-called "transverse forces", which have been used to accelerate the sampling of potential energy surfaces (Ghimenti et al., 2023) and stabilize self-propelling particles (Du & Vaikuntanathan, 2024). These forces are easy to implement in general. These physics-motivated insights can potentially be incorporated in state-of-the-art neural networks, for example, as part of an energy-based LSTM (Hoover et al., 2024) to construct associative memory based temporal pattern recognition networks.

## 2 INTRODUCING A DEGREE OF FREEDOM TO ACCOUNT FOR DELAYED INTEGRATION

The model of Kanter & Sompolinsky (1986) extends the Hopfield model to the retrieval of a sequence of patterns. The firing activity of neuron $i = 1, \ldots, N$ is represented by a discrete spin $\sigma_i \in \{\pm 1\}$ obeying the dynamics

$$\sigma_i(t+1) = \mathrm{sgn}(h_i(t)) \tag{1}$$

$$h_i(t) = \frac{\partial \mathcal{H}}{\partial \sigma_i} + \frac{\lambda}{N} \sum_{j,\mu} \xi_i^{\mu+1} \xi_j^{\mu} \int_{-\infty}^{t} dt' w(t-t') \sigma_j(t') \tag{2}$$

$$\frac{\partial \mathcal{H}}{\partial \sigma_i} = \frac{1}{N} \sum_{j,\mu} \xi_i^{\mu} \xi_j^{\mu} \sigma_j \tag{3}$$

where $\{\boldsymbol{\xi}^{\mu}\}_{\mu=1}^{p}$ is the sequence of $p$ patterns to be retrieved, and the pattern elements are independently and identically distributed (i.i.d.) random variables chosen as $\pm 1$ with equal probability. The coupling between each pair of spins $i$ and $j$ consists of a symmetric part (equation 2, first term, which is defined in equation 3) and an asymmetric part (equation 2, second term). In particular, the asymmetric coupling is determined by strength $\lambda$ and an asymmetric matrix whose $ij$th element is $\sum_{\mu} \xi_i^{\mu+1} \xi_j^{\mu}$, which helps the system to evolve from pattern "$\mu$" to pattern "$\mu + 1$". The integrating kernel, $w(t - t')$, ensures coherent oscillations. Only when the state has aligned with pattern $\mu$ for a characteristic time, $\tau$, it moves to the next pattern. Some forms of $w(t)$ are given by: $\frac{1}{\tau} \exp(-\frac{t}{\tau})$; and $\frac{1}{\tau}$ for t $< \tau$, 0 for t $> \tau$. In general, $w(t)$ needs to be a decreasing function and satisfy $\int_0^{\infty} w(t) dt = 1$.

We find that to obtain the same result for continuous spin dynamics would be to add a new variable ($\boldsymbol{x}$) that takes care of the integration kernel in the evolution equation:

$$\frac{\partial \sigma_i}{\partial t} = -u(t)\sigma_i - \frac{\partial \mathcal{H}}{\partial \sigma_i} + \lambda \xi_i^{\mu+1} x^{\mu} (1 + (x^{\mu})^2) + \eta_i(t) \tag{4}$$

$$\tau \frac{\partial x^{\mu}}{\partial t} = -f(x^{\mu}) x^{\mu} + \frac{1}{N} \xi_i^{\mu} \sigma_i \tag{5}$$

$$\frac{\partial \mathcal{H}}{\partial \sigma_i} = \xi_i^{\mu} \frac{1}{N} \sum_j \xi_j^{\mu} \sigma_j + u_0 \xi_i^{\mu} \left( \frac{1}{N} \sum_j \xi_j^{\mu} \sigma_j \right)^3 \tag{6}$$

One can view the expanded system as a 2-layer neural network, where $\boldsymbol{\sigma}$ represents the visible neurons and $\boldsymbol{x}$ represents the hidden neurons. Here, $u(t)$ enforces the L2 regularization $\|\boldsymbol{\sigma}\|^2 = N$ at all times. For such "spherical models", the quartic terms are essential to have pattern retrieval (Crisanti & Sompolinsky, 1987; Bollé et al., 2003). $\eta_i$ represents the thermal fluctuations in the system with $\langle \eta_i \rangle = 0$ and $\langle \eta_i(t) \eta_j(t') \rangle = 2T\delta_{ij}\delta(t - t')$. If $f(x^{\mu}) = 1$, then one recovers the cycling equation with $w(t) = \tau^{-1} \exp(-\frac{t}{\tau})$ (with an additional quartic term). Similarly, for other choices of the function, $f$, one can obtain various forms for $w(t)$. From the perspective of signal processing, the ability to realize the integrating kernel $w(t)$ using an auxiliary variable $\boldsymbol{x}$ follows immediately from the fact that $w(t)$ acts as a first-order low-pass filter (Tan & Jiang, 2018) in equation 2. Fig. 1 depicts pattern cycling with the additional degree of freedom. We note that Lu & Wu (2024) introduced hidden neurons for sequence retrieval but the connection to Kanter & Sompolinsky (1986) was not made apparent. In the absence of hidden neurons, one needs to store the entire history of the system for performing the convolution operation. Adding a new hidden variable alleviates this storage cost.

## 3 ENHANCING THE PATTERN CYCLING WITH A TRANSVERSE FORCE

In this section, we apply a transverse force to a continuous version of a modern Hopfield network (Krotov & Hopfield, 2016) whose asymmetric interactions induces pattern cycling (Chaudhry et al., 2024). The system evolves as (Fig. 2a)

$$\frac{d\boldsymbol{\sigma}}{dt} = \mathbf{F}_{\text{cyc}} + \mathbf{F}_{\text{sph}} + \mathbf{F}_{\perp}(\boldsymbol{\sigma}), \tag{7}$$

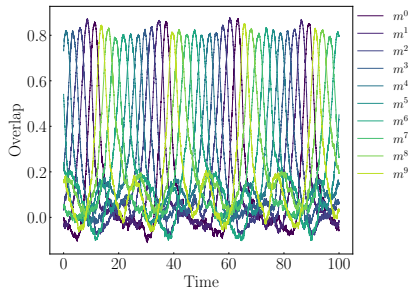


Figure 1: Pattern cycling with the help of asymmetric interaction with the additional degree of freedom. The curves denote the traces for the overlaps with the 10 patterns stored in the system. For the purposes of this simulation, the following parameters were chosen: $N = 400, \alpha = 0.025, T = 0.1, \lambda = 1, \tau = 2$. The pattern cycling was carried out by integrating Eqs. 4, 5 and 6 for $2 \times 10^5$ steps with $\Delta t = 10^{-3}$.

where

$$\mathbf{F}_{\text{cyc}} = \sum_{\mu=1}^{p} \boldsymbol{\xi}^{\mu+1} \phi(m^\mu) \tag{8}$$

drives the cycling between patterns,

$$\mathbf{F}_{\text{sph}} = -\nabla_{\boldsymbol{\sigma}} U(\|\boldsymbol{\sigma}\|^2 - N) \tag{9}$$

implements the L2 regularization $\|\boldsymbol{\sigma}\|^2 \approx N$, and the transverse force (Ghimenti et al., 2023; Du & Vaikuntanathan, 2024)

$$\mathbf{F}_{\perp} = \mathbf{A}(\mathbf{F}_{\text{cyc}} + \mathbf{F}_{\text{sph}}) \tag{10}$$

is perpendicular to $\mathbf{F}$ since $\mathbf{A}$ is antisymmetric. Here, the pattern elements $\xi_i^\mu \sim \mathcal{N}(0,1)$ are i.i.d., $\phi(y)$ is a monotonically increasing nonlinear function, $m^\mu = \frac{1}{N}(\boldsymbol{\xi}^\mu)^T \sigma_i$ is the overlap of the spins with pattern $\mu$, and the specific form of $U(\|\boldsymbol{\sigma}\|^2 - N)$—a convex function minimized when $\|\boldsymbol{\sigma}\|^2 = N$—depends on the choice of $\phi$. The elements of $\mathbf{A}$ are i.i.d. as $A_{ij} = -A_{ji} \sim \mathcal{N}(0, a^2/N)$ (Singh et al., 1995; Chengxiang et al., 2000), where $a$ is the strength of the transverse force.

Figs. 2b-2c show cycling between $p = 10$ patterns for $\phi(y) = y^3$ and $U(\|\boldsymbol{\sigma}\|^2 - N) = \frac{u}{4}(\|\boldsymbol{\sigma}\|^2 - N)^2$, where $u = 2$ is the strength of the L2 regularization. We see that the transverse force speeds up the pattern cycling. In addition, the transverse force sharpens the transitions between patterns: the overlap with one pattern decays more before the overlap with the next pattern peaks.

To theoretically understand this enhancement of pattern cycling, we consider the limit of weak transverse force, such that we can truncate the dynamics of the pattern overlaps beyond $O(\mathbf{A}^2)$. We also assume the thermodynamic limit of spins, $N \to \infty$, and a finite number $p$ of stored patterns. Starting from pattern 1, at sufficiently short times $t$, the overlap with pattern 2 evolves as (Appendix D)

$$\frac{dm^2}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|} U}{\|\boldsymbol{\sigma}\|} \left[ m^2 - \frac{t}{N} \phi(m^1) \left\| \mathbf{A}\boldsymbol{\xi}^2 \right\|^2 \right] + \frac{1}{N} \phi(m^1) \left\| \boldsymbol{\xi}^2 \right\|^2, \tag{11}$$

which shows that the transverse force ($\mathbf{A}$) suppresses the decay contribution from the L2 regularization ($U$). Thus, the overlap with $\boldsymbol{\xi}^2$ rises faster, in agreement with the simulations (Figs. 2d-2e).

Based on this perturbative analysis, we arrive at the following intuitive picture of the accelerated pattern cycling (Figs. 2d-2e). Without the transverse force, as the spins transition from $\boldsymbol{\xi}^1$ to $\boldsymbol{\xi}^2$, the L2 regularization restrains the magnitude of the spin vector by exerting a force in the opposite direction, which has some component along $-\boldsymbol{\xi}^2$. With the transverse force, the spins acquire a component along $\mathbf{A}\boldsymbol{\xi}^2$ as they move towards $\boldsymbol{\xi}^2$. As a result, the L2 regularization counters with a force having a component along $-\mathbf{A}\boldsymbol{\xi}^2$, and the corresponding transverse component is in the direction of $\mathbf{A}(-\mathbf{A}\boldsymbol{\xi}^2) = -\mathbf{A}^2\boldsymbol{\xi}^2$. Since $\mathbf{A}$ is antisymmetric, then $\mathbf{A}^2$ is negative semidefinite, and so $-\mathbf{A}^2\boldsymbol{\xi}^2 \parallel -(-\boldsymbol{\xi}^2)$. Thus, the transverse component along $-\mathbf{A}^2\boldsymbol{\xi}^2$ goes against the L2 regularization, allowing the spins to move faster towards $\boldsymbol{\xi}^2$.

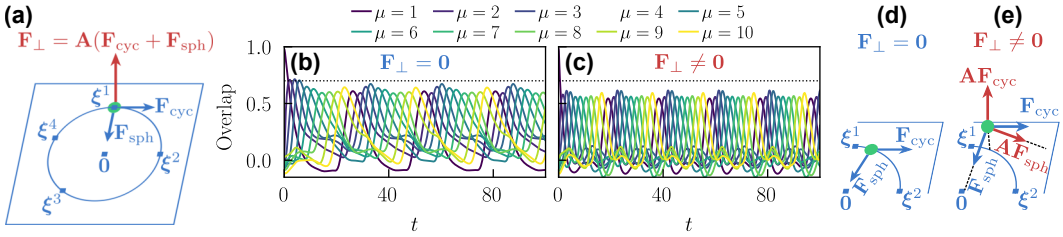


Figure 2: Pattern cycling enhanced by a transverse force. (a) Schematic illustration of the model. (b-c) Dynamics of the overlaps with pattern $\mu$ (b) without and (c) with ($a = 2$) the transverse force. We numerically integrate equation 7 with time step $\Delta t = 10^{-3}$. The black dotted line is a simply a guide to the eye. (d-e) Schematic illustration of the forces after the system evolves for a short time (d) without and (e) with a weak transverse force.

## 4 Discussion

Drawing inspiration from nonequilibrium physics, we have presented new representations of temporal sequence retrieval in the Hopfield model. We showed how the memory kernel employed by (Kanter & Sompolinsky, 1986) to facilitate cycling between patterns can be alternatively realized using hidden neurons. We also demonstrated that transverse forces can speed up and sharpen the transitions between patterns.

Although we considered a particular model of sequence retrieval, we expect our results to be fairly general. For simplicity, we considered here a transverse force determined by an antisymmetric matrix $\mathbf{A}$, whose element are i.i.d. Gaussian (equation 10); nevertheless, since $\mathbf{A}$ was generated independently of the sequence to be retrieved, other choices of antisymmetric $\mathbf{A}$ should be suitable as well. While we focused on models with L2 regularization, our results do not crucially depend on a certain type of regularization and thus are expected to hold across a broad range of neural networks. Our results on transverse forces might also be relevant to studies [see Farrell & Pehlevan (2024) and references therein] on how biological neural networks learn to generate sequential activity at a particular speed. Given the focus here on an RNN (i.e., a Hopfield model) that evolves continuously in time, we anticipate that our proposed strategies for sequence retrieval could be inapplicable in cases where the RNN evolves discretely in time. Experiments on standard sequence modeling tasks should be done to start evaluating the generality of our work.

Our physics-inspired findings can be implemented in state-of-the-art neural networks through the use of Hopfield submodules (Hoover et al., 2024). For instance, we envision supplementing an LSTM with a Hopfield network, which can feature memory kernel-like hidden neurons and transverse forces. The trainable parameters of the hybrid neural network could include, for example, the strength of the coupling to hidden neurons and the form of the transverse forces.

### Acknowledgments

MD and SV are supported by DOE BES, Grant No. DESC0019765.

## References

D. Bollé, Th M. Nieuwenhuizen, I. Pérez Castillo, and T. Verbeiren. A spherical Hopfield model. *Journal of Physics A: Mathematical and General*, 36(41):10269–10277, October 2003.

Hamza Chaudhry, Jacob Zavatone-Veth, Dmitry Krotov, and Cengiz Pehlevan. Long sequence hopfield memory. *Advances in Neural Information Processing Systems*, 36, 2024.

Zhang Chengxiang, Chandan Dasgupta, and Manoranjan P. Singh. Retrieval Properties of a Hopfield Model with Random Asymmetric Interactions. *Neural Computation*, 12(4):865–880, April 2000.

A. Crisanti and H. Sompolinsky. Dynamics of spin systems with randomly asymmetric bonds: Langevin dynamics and a spherical model. *Physical Review A*, 36(10):4922–4939, November 1987.

Matthew Du and Suriyanarayanan Vaikuntanathan. Hidden nonreciprocity as a stabilizing effective potential in active matter, January 2024. arXiv:2401.14690 [cond-mat].

Matthew Farrell and Cengiz 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):e2309876121, August 2024. doi: 10.1073/pnas.2309876121. URL https://www.pnas.org/doi/abs/10.1073/pnas.2309876121.

Federico Ghimenti, Ludovic Berthier, Grzegorz Szamel, and Frédéric van Wijland. Sampling Efficiency of Transverse Forces in Dense Liquids. *Physical Review Letters*, 131(25):257101, December 2023.

Benjamin Hoover, Yuchen Liang, Bao Pham, Rameswar Panda, Hendrik Strobelt, Duen Horng Chau, Mohammed Zaki, and Dmitry Krotov. Energy transformer. *Advances in Neural Information Processing Systems*, 36, 2024.

John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

Ido Kanter and Haim Sompolinsky. Temporal association in asymmetric neural networks. *Physical Review Letters*, 57(22):2861–2864, 1986. doi: 10.1103/PhysRevLett.57.2861.

Dmitry Krotov and John J. Hopfield. Dense associative memory for pattern recognition. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper/2016/file/eaae339c4d89fc102edd9dbdb6a28915-Paper.pdf.

Yao Lu and Si Wu. Learning sequence attractors in recurrent networks with hidden neurons, 2024. URL https://arxiv.org/abs/2404.02729.

Christian K Machens, Ranulfo Romo, and Carlos D Brody. Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science*, 307(5712):1121–1124, 2005.

Manoranjan P. Singh, Zhang Chengxiang, and Chandan Dasgupta. Fixed points in a Hopfield model with random asymmetric interactions. *Physical Review E*, 52(5):5261–5272, November 1995.

Li Tan and Jean Jiang. *Digital signal processing: fundamentals and applications*. Academic press, 2018.

Han Yan, Lei Zhao, Liang Hu, Xidi Wang, Erkang Wang, and Jin Wang. Nonequilibrium landscape theory of neural networks. *Proceedings of the National Academy of Sciences*, 110(45):E4185–E4194, November 2013.

## A EXACT DYNAMICS OF PATTERN OVERLAPS: NO TRANSVERSE FORCE

Without the transverse force (equation 7-equation 9, $\mathbf{F}_\perp = \mathbf{0}$), the dynamics of the pattern overlaps is exactly given by

$$\frac{dm^\mu}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|} U}{\|\boldsymbol{\sigma}\|} m^\mu + \frac{1}{N} \sum_{\nu=1}^{p} \left(\boldsymbol{\xi}^\mu\right)^T \boldsymbol{\xi}^{\nu+1} \phi(m^\nu), \tag{12}$$

$$\frac{d\|\boldsymbol{\sigma}\|}{dt} = -\partial_{\|\boldsymbol{\sigma}\|} U + \frac{N}{\|\boldsymbol{\sigma}\|} \sum_{\nu=1}^{p} m^{\nu+1} \phi(m_\nu). \tag{13}$$

## B EXACT DYNAMICS OF PATTERN OVERLAPS: TRANSVERSE FORCE

With the transverse force (equation 7-equation 10, $\mathbf{F}_\perp \neq \mathbf{0}$), the dynamics of the pattern overlaps is exactly described by a hierarchy of equations of motion,

$$\frac{dm^\mu}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|}U}{\|\boldsymbol{\sigma}\|}\left(m^\mu - \tilde{m}^\mu_{(1)}\right) + \frac{1}{N}\sum_{\nu=1}^{p}\left(\boldsymbol{\xi}^\mu\right)^T(\mathbf{I}+\mathbf{A})\boldsymbol{\xi}^{\nu+1}\phi(m^\nu), \tag{14}$$

$$\frac{d\|\boldsymbol{\sigma}\|}{dt} = -\partial_{\|\boldsymbol{\sigma}\|}U + \frac{N}{\|\boldsymbol{\sigma}\|}\sum_{\nu=1}^{p}\left(m^{\nu+1} + \tilde{m}^{\nu+1}_{(1)}\right)\phi(m_\nu), \tag{15}$$

$$\frac{d\tilde{m}^\mu_{(n)}}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|}U}{\|\boldsymbol{\sigma}\|}\left(\tilde{m}^\mu_{(n)} - \tilde{m}^\mu_{(n+1)}\right) + \frac{1}{N}\sum_{\nu=1}^{p}\left(\mathbf{A}^n\boldsymbol{\xi}^\mu\right)^T(\mathbf{I}+\mathbf{A})\boldsymbol{\xi}^{\nu+1}\phi(m^\nu), \quad n = 1, 2, \ldots \tag{16}$$

where $\tilde{m}^\mu_{(n)} = \frac{1}{N}\left(\mathbf{A}^n\boldsymbol{\xi}^\mu\right)^T\boldsymbol{\sigma}$ is the overlap with the $n$-fold antisymmetrically transformed patterns $\mathbf{A}^n\boldsymbol{\xi}^\mu$. In this exact evolution for the pattern overlaps, the degrees of freedom include not only the pattern overlaps themselves and the norm of the spin vector, as in the case without a transverse force (equation 12-equation 13), but also the overlaps $\tilde{m}^\mu_{(n)} = \frac{1}{N}\left(\mathbf{A}^n\boldsymbol{\xi}^\mu\right)^T\boldsymbol{\sigma}$ with the $n$-fold antisymmetrically transformed patterns $\mathbf{A}^n\boldsymbol{\xi}^\mu$.

## C TRUNCATING DYNAMICS OF PATTERN OVERLAPS TO ORDER $n$ IN TRANSVERSE FORCE

In principle, one can obtain the numerically exact dynamics from equation 14-equation 16 by keeping terms up to $O(\mathbf{A}^n)$ for sufficiently large $n$. Specifically, we would retain all equations up to those for $\{m^\mu_{(n)}\}$, where the equations of motion for these highest-order quantities are themselves truncated according to

$$\frac{d\tilde{m}^\mu_{(n)}}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|}U}{\|\boldsymbol{\sigma}\|}\tilde{m}^\mu_{(n)} + \frac{1}{N}\sum_{\nu=1}^{p}\left(\mathbf{A}^n\boldsymbol{\xi}^\mu\right)^T\boldsymbol{\xi}^{\nu+1}\phi(m^\nu) + O(\mathbf{A}^{n+1}). \tag{17}$$

## D APPROXIMATE DYNAMICS OF PATTERN OVERLAPS: WEAK TRANSVERSE FORCE AND FINITE NUMBER OF STORED PATTERNS IN THE THERMODYNAMIC LIMIT

Here, we consider the dynamics of the pattern overlaps (equation 14-equation 16) assuming the thermodynamic limit of spins, a finite number of stored patterns, and a weak transverse force. Specifically, we assume $N \to \infty$, finite $p$, and $A_{ij}$ values small enough to approximate equation 14-equation 16 to $O(\mathbf{A})^2$, respectively. By examining the short-time dynamics initialized at a stored pattern configuration, we gain intuition for why the transverse force speeds up pattern cycling.

Up to $O(\mathbf{A}^2)$, the dynamics of equation 14-equation 16 is given by (equation 17)

$$\frac{dm^\mu}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|}U}{\|\boldsymbol{\sigma}\|}\left(m^\mu - \tilde{m}^\mu_{(1)}\right) + \frac{1}{N}\sum_{\nu=1}^{p}\left(\boldsymbol{\xi}^\mu\right)^T(\mathbf{I}+\mathbf{A})\boldsymbol{\xi}^{\nu+1}\phi(m^\nu), \tag{18}$$

$$\frac{d\|\boldsymbol{\sigma}\|}{dt} = -\partial_{\|\boldsymbol{\sigma}\|}U + \frac{N}{\|\boldsymbol{\sigma}\|}\sum_{\nu=1}^{p}\left(m^{\nu+1} + \tilde{m}^{\nu+1}_{(1)}\right)\phi(m_\nu), \tag{19}$$

$$\frac{d\tilde{m}^\mu_{(1)}}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|}U}{\|\boldsymbol{\sigma}\|}\left(\tilde{m}^\mu_{(1)} - \tilde{m}^\mu_{(2)}\right) + \frac{1}{N}\sum_{\nu=1}^{p}\left(\mathbf{A}\boldsymbol{\xi}^\mu\right)^T(\mathbf{I}+\mathbf{A})\boldsymbol{\xi}^{\nu+1}\phi(m^\nu), \tag{20}$$

$$\frac{d\tilde{m}^\mu_{(2)}}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|}U}{\|\boldsymbol{\sigma}\|}\tilde{m}^\mu_{(2)} + \frac{1}{N}\sum_{\nu=1}^{p}\left(\mathbf{A}^2\boldsymbol{\xi}^\mu\right)^T\boldsymbol{\xi}^{\nu+1}\phi(m^\nu). \tag{21}$$

To further reduce the equations of motion, we hereafter assume $N \to \infty$. Using the facts that the pattern elements are i.i.d. random variables $\xi_i^\mu \sim \mathcal{N}(0,1)$ and the elements of the antisymmetric matrix $\mathbf{A}$ are i.i.d. random variables $A_{ij} = -A_{ji} \sim \mathcal{N}(0, a^2/N)$, it is straightforward (but tedious) to compute the following averages over $\{\boldsymbol{\xi}^\mu\}$ and $\mathbf{A}$:

$$\left\langle \left[ \frac{1}{N} \left(\boldsymbol{\xi}^\mu\right)^T \boldsymbol{\xi}^\nu \right]^2 \right\rangle = \begin{cases} 1 + O(N^{-1}) & \text{if } \mu = \nu, \\ O(N^{-1}) & \text{if } \mu \neq \nu, \end{cases} \tag{22}$$

$$\left\langle \left[ \frac{1}{N} \left(\boldsymbol{\xi}^\mu\right)^T \mathbf{A}\boldsymbol{\xi}^\nu \right]^2 \right\rangle = \begin{cases} 0 & \text{if } \mu = \nu, \\ O(N^{-1}) & \text{if } \mu \neq \nu, \end{cases} \tag{23}$$

$$\left\langle \left[ \frac{1}{N} \left(\boldsymbol{\xi}^\mu\right)^T \mathbf{A}^2\boldsymbol{\xi}^\nu \right]^2 \right\rangle = \begin{cases} a^4 + O(N^{-1}) & \text{if } \mu = \nu, \\ O(N^{-1}) & \text{if } \mu \neq \nu. \end{cases} \tag{24}$$

Thus,

$$\frac{1}{N} \left(\boldsymbol{\xi}^\mu\right)^T \boldsymbol{\xi}^\nu = \begin{cases} O(N^0) & \text{if } \mu = \nu, \\ O(N^{-1/2}) & \text{if } \mu \neq \nu, \end{cases} \tag{25}$$

$$\frac{1}{N} \left(\boldsymbol{\xi}^\mu\right)^T \mathbf{A}\boldsymbol{\xi}^\nu = \begin{cases} 0 & \text{if } \mu = \nu, \\ O(N^{-1/2}) & \text{if } \mu \neq \nu, \end{cases} \tag{26}$$

$$\frac{1}{N} \left(\boldsymbol{\xi}^\mu\right)^T \mathbf{A}^2\boldsymbol{\xi}^\nu = \begin{cases} O(N^0) & \text{if } \mu = \nu, \\ O(N^{-1/2}) & \text{if } \mu \neq \nu. \end{cases} \tag{27}$$

Using these scalings and also assuming a finite number of patterns, we can neglect the contributions to equation 18-equation 21 that vanish as $N \to \infty$ at all times $t$, resulting in

$$\frac{dm^\mu}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|} U}{\|\boldsymbol{\sigma}\|} \left( m^\mu - \tilde{m}^\mu_{(1)} \right) + \frac{1}{N} \left\|\boldsymbol{\xi}^\mu\right\|^2 \phi(m^{\mu-1}), \tag{28}$$

$$\frac{d\|\boldsymbol{\sigma}\|}{dt} = -\partial_{\|\boldsymbol{\sigma}\|} U + \frac{N}{\|\boldsymbol{\sigma}\|} \sum_{\nu=1}^{p} \left( m^{\nu+1} + \tilde{m}^{\nu+1}_{(1)} \right) \phi(m^{\mu-1}), \tag{29}$$

$$\frac{d\tilde{m}^\mu_{(1)}}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|} U}{\|\boldsymbol{\sigma}\|} \left( \tilde{m}^\mu_{(1)} - \tilde{m}^\mu_{(2)} \right) + \frac{1}{N} \left\|\mathbf{A}\boldsymbol{\xi}^\mu\right\|^2 \phi(m^{\mu-1}), \tag{30}$$

$$\frac{d\tilde{m}^\mu_{(2)}}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|} U}{\|\boldsymbol{\sigma}\|} \tilde{m}^\mu_{(2)} - \frac{1}{N} \left\|\mathbf{A}\boldsymbol{\xi}^\mu\right\|^2 \phi(m^{\mu-1}). \tag{31}$$

We now consider the dynamics starting from pattern 1, $\boldsymbol{\sigma}(0) = \frac{\sqrt{N}}{\|\boldsymbol{\xi}^1\|}\boldsymbol{\xi}^1$, at short times $t$,

$$\frac{dm^1}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|} U}{\|\boldsymbol{\sigma}\|} \left( m^1 - \tilde{m}^{(1)}_1 \right), \tag{32}$$

$$\frac{dm^2}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|} U}{\|\boldsymbol{\sigma}\|} \left( m^2 - \tilde{m}^2_{(1)} \right) + \frac{1}{N} \phi(m^1) \left\|\boldsymbol{\xi}^2\right\|^2, \tag{33}$$

$$\frac{d\|\boldsymbol{\sigma}\|}{dt} = -\partial_{\|\boldsymbol{\sigma}\|} U + \frac{N}{\|\boldsymbol{\sigma}\|} \phi(m^1) \left( m_2 + \tilde{m}^2_{(1)} \right), \tag{34}$$

$$\frac{d\tilde{m}^1_{(1)}}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|} U}{\|\boldsymbol{\sigma}\|} \left( \tilde{m}^1_{(1)} - \tilde{m}^1_{(2)} \right), \tag{35}$$

$$\frac{d\tilde{m}^1_{(2)}}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|} U}{\|\boldsymbol{\sigma}\|} \tilde{m}^1_{(2)}, \tag{36}$$

$$\frac{d\tilde{m}^2_{(1)}}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|} U}{\|\boldsymbol{\sigma}\|} \left( \tilde{m}^2_{(1)} - \tilde{m}^2_{(2)} \right) + \frac{1}{N} \phi(m^1) \left\|\mathbf{A}\boldsymbol{\xi}^2\right\|^2, \tag{37}$$

$$\frac{d\tilde{m}^2_{(2)}}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|} U}{\|\boldsymbol{\sigma}\|} \tilde{m}^2_{(2)} - \frac{1}{N} \phi(m^1) \left\|\mathbf{A}\boldsymbol{\xi}^2\right\|^2. \tag{38}$$

The corresponding quantities for the other patterns, $\mu > 2$, remain 0 at short times in the thermodynamic limit. Next, we apply the approximation $\frac{dy}{dt} \approx \frac{y(t)-y(0)}{t}$ for general $y(t)$ to equation 35, equation 36, equation 37, and equation 38 and subsequently solve for $\tilde{m}^1_{(1)}(t)$, $\tilde{m}^1_{(2)}(t)$, $\tilde{m}^2_{(1)}(t)$, and $\tilde{m}^2_{(1)}(t)$, respectively. Plugging these solutions into equation 32-equation 34 yields

$$\frac{dm^1}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|}U}{\|\boldsymbol{\sigma}\|}\left[m^1 - kt\tilde{m}^1_{(2)}(0)\right], \tag{39}$$

$$\frac{dm^2}{dt} = -\frac{\partial_{\|\boldsymbol{\sigma}\|}U}{\|\boldsymbol{\sigma}\|}\left[m^2 - \frac{t}{N}\phi(m^1)\left\|\mathbf{A}\boldsymbol{\xi}^2\right\|^2\right] + \frac{1}{N}\phi(m^1)\left\|\boldsymbol{\xi}^2\right\|^2, \tag{40}$$

$$\frac{d\|\boldsymbol{\sigma}\|}{dt} = -\partial_{\|\boldsymbol{\sigma}\|}U + \frac{N}{\|\boldsymbol{\sigma}\|}\phi(m^1)\left[m^2 + \frac{t}{N}\phi(m^1)\left\|\mathbf{A}\boldsymbol{\xi}^2\right\|^2\right], \tag{41}$$

where we have kept terms up to $O(t)$ on the right-hand side of each equation. In the equation for quantity $y = m^1, m^2, \|\boldsymbol{\sigma}\|$, the terms containing $-\partial_{\|\boldsymbol{\sigma}\|}U$ contribute to the decay of $y$, while the remaining terms contribute to the growth of $y$. With this interpretation, we can understand how the transverse force ($\mathbf{A} \neq \mathbf{0}$) affects the dynamics of $m_1$, $m_2$, and $\|\boldsymbol{\sigma}\|$. From equation 39, we see that the transverse force speeds up the decay of $m^1$, since $\tilde{m}^1_{(2)}(0) = (\boldsymbol{\xi}^1)^T\mathbf{A}^2\boldsymbol{\xi}^1/\|\boldsymbol{\xi}^1\|^2 < 0$ and $\mathbf{A}^2$ is a negative semidefinite matrix for antisymmetric $\mathbf{A}$. In contrast, equation 40 shows that the transverse force speeds up the (overall) growth of $m^2$ by suppressing the decay contribution. Similarly, equation 41 shows that the transverse force speeds up the growth of the norm of the spin vector.