Transient dynamics of associative memory models

David G. Clark [*,†]
Zuckerman Institute, Columbia University, New York, New York 10027, USA
and Kavli Institute for Brain Science, Columbia University, New York, New York 10027, USA

(Received 27 October 2025; accepted 13 April 2026; published 6 May 2026)

Associative memory models such as the Hopfield network and its dense generalizations with higher-order interactions exhibit a "blackout catastrophe"—a discontinuous transition where stable memory states abruptly vanish when the number of stored patterns exceeds a critical capacity. This transition is often interpreted as rendering networks unusable beyond capacity limits. We argue that this interpretation is largely an artifact of the equilibrium perspective. We derive dynamical mean-field equations for graded-activity dense associative memory models, with the Hopfield model as a special case, using a bipartite cavity approach. We solve the resulting self-consistent equations using an iterative numerical scheme. We show that patterns can be transiently retrieved with high accuracy above capacity despite the absence of stable attractors. This occurs because slow regions persist in the above-capacity energy landscape near stored patterns as lingering traces of the stable basins that existed below capacity. The same transient-retrieval effect occurs in below-capacity networks initialized outside basins of attraction. "Transient-recovery curves" provide a concise visual summary of these effects, revealing graceful, noncatastrophic changes in retrieval behavior above capacity and allowing us to compare the behavior across interaction orders. This dynamical perspective reveals energy landscape structure obscured by equilibrium analysis, including slow regions near stored patterns that persist above capacity, and suggests biological neural circuits may exploit transient dynamics for memory retrieval. Furthermore, our approach suggests ways of understanding computational properties of neural circuits without reference to fixed points and yields new theoretical results on generalizations of the Hopfield model.

DOI: 10.1103/42y2-bsh1

I. INTRODUCTION

The Hopfield model is a recurrent neural network with weights constructed through a Hebbian learning rule that can store and retrieve patterns, therefore functioning as a memory device [1,2]. Its dynamics are governed by an energy function, permitting its analysis within equilibrium statistical mechanics, in particular, using methods for disordered systems such as the replica method [3,4].

A key result from this line of work is that the standard Hopfield model can successfully store and retrieve $P = O(N)$ random patterns as stable fixed points, where N is the number of neurons. Beyond this capacity, interference among patterns encoded in the connectivity destroys these stable states [3,4], a phenomenon known as "blackout catastrophe" [5,6]. This represents a discontinuous, first-order phase transition: the overlap between network activity and a target pattern remains high, corresponding to memory retrieval, until a critical capacity of $P \approx 0.14N$ for binary-spin models. Beyond this capacity, the high-overlap solution vanishes, and only the zero-overlap solution remains.

The limited capacity of the Hopfield model has motivated various generalizations. The dense associative memory model, introduced and studied decades ago [7–10] and connected to modern deep learning by Krotov and Hopfield [11,12], achieves capacity $P = O(N^n)$ with $(n+1)$-way neuronal interactions. This represents a qualitative improvement over the Hopfield model's $P = O(N)$ capacity for $n > 1$, with the Hopfield model corresponding to the special case $n = 1$ with pairwise interactions. These generalized models thus liberate storage capacity from pattern dimensionality, to which storage capacity is constrained to be proportional in the Hopfield case. A caveat to this line of work is that while the model achieves $P = O(N^n)$ capacity, it can equivalently be formulated as a bipartite system with $P + N$ units using pairwise interactions, thus recovering linear scaling in the total number of units. Rather than appealing to this bipartite formulation, some authors have proposed biological implementations of effectively higher-order neuronal interactions, although these remain speculative [13]. Regardless of how significant one finds the increased capacity, what remains interesting and nontrivial are the pattern retrieval dynamics, the focus of this work, that allow these models to recall stored patterns from partial or corrupted inputs within densely packed feature spaces.

Like the Hopfield model, these generalized models possess energy functions governing their dynamics and exhibit discontinuous vanishing of stable retrieval states when capacity

[*]Present address: Kempner Institute for the Study of Natural and Artificial Intelligence, Harvard University, Cambridge, Massachusetts 02138, USA.
[†]Contact author: dgclark@fas.harvard.edu

is exceeded [7], although this is less limiting in practice given their increased capacity for $n > 1$.

The capacity constraints that seemingly render associative memory models unusable beyond their capacity limits are derived from equilibrium analyses that probe stable memory states; that is, energy landscape local minima, but provide limited insight into the system's behavior during transient evolution. We therefore adopt a dynamical rather than an equilibrium perspective on associative memory models. We study the out-of-equilibrium, transient dynamics of these systems and demonstrate that the blackout catastrophe is not catastrophic when viewed dynamically. In particular, even when stable fixed points no longer exist beyond the critical capacity, memories can still be transiently recalled, often with high accuracy, during evolution from initial conditions. Analysis of the energy function reveals that this transient retrieval reflects the persistence of slow regions in the energy landscape near stored patterns, even above capacity.

Neural circuit dynamics are frequently characterized through fixed points and stability analysis. Indeed, constructing a "dynamical skeleton" based on fixed points and transitions between them has been among the most successful approaches for understanding artificial [14] and biological recurrent neural circuits [15]. However, transient dynamics outside of fixed points are likely crucial for neural computation and require new analysis methods [16]. For memory systems, we propose "transient-recovery curves," which characterize memory retrieval performance without requiring stable attractor states. These curves reveal a graceful degradation of retrieval performance as capacity is exceeded, rather than abrupt failure. By varying the number of stored patterns, these curves sweep out a family that can be compared across different interaction orders n, permitting analysis of how higher-order interactions shape retrieval dynamics.

To study these transient dynamics in the $N \to \infty$ limit, we develop a dynamical mean-field theory (DMFT) for graded-activity dense associative memory models storing an appropriately scaled infinite number of patterns. We solve the resulting self-consistent equations using iterative numerical methods similar to those described by Roy *et al.* [17] for ecological systems. The equilibrium statistical mechanics of graded-activity Hopfield models (the $n = 1$ case) was analyzed by Kühn *et al.* [18]. While several works from the 1980s and 1990s derived DMFT equations for Hopfield models, they typically studied binary spins in discrete time rather than continuous variables in continuous time and, crucially, could not numerically solve the self-consistent equations. The binary-spin restriction limits relevance to machine learning, which requires differentiability. We review this historical context in detail in the Discussion. The iterative scheme we use here is made feasible by modern computational resources, particularly GPU acceleration. Our approach allows us to capture the full temporal evolution of these systems, reveals rich transient dynamics that were previously inaccessible to theoretical analysis, and extends all of these analyses to higher-order generalizations of the Hopfield model.

II. HOPFIELD AND DENSE ASSOCIATIVE MEMORY MODELS

We now define the class of models considered in this paper: the dense associative memory model for arbitrary n, with the Hopfield model corresponding to $n = 1$. This can be done through either a neuronal [Fig. 1(a)] or a bipartite [Fig. 1(b)] formulation, the latter of which is amenable to a cavity analysis [Fig. 1(c)]. We then specify a generative process for the stored patterns and initial conditions to enable a large-N analysis. Finally, we distinguish between condensed and uncondensed patterns and show through a signal-to-noise argument that the capacity scales as $P = O(N^n)$.

A. Neuronal formulation

The dense associative memory model generalizes the Hopfield model by introducing higher-order, $(n+1)$-way interactions among neurons [Fig. 1(a); note that prior works typically define n to be the order of interactions itself; that is, what we call $n+1$]. Consider N neurons with preactivations $x_i(t)$ and nonlinearly transformed activations $\phi_i(t) = \phi(x_i(t))$, where $i \in \{1, 2, \ldots, N\}$ indexes neurons and $\phi(x)$ is a bounded and monotonic neuronal nonlinearity. The neuronal dynamics are governed by

$$x_i(t) = (1 - \Delta t)x_i(t-1) + \Delta t \left[\frac{g}{\sqrt{\alpha}} \sum_{j_1, j_2, \cdots, j_n} T_{i j_1 j_2 \cdots j_n} \phi_{j_1}(t-1)\phi_{j_2}(t-1)\cdots\phi_{j_n}(t-1) + I_i(t-1) \right], \tag{1}$$

where $t \in \{1, 2, \ldots, T\}$ indexes discrete time steps, Δt is the time-step size, and $I_i(t)$ are external inputs that serve as source terms in the mean-field analysis (they can be set to zero when not needed for this purpose). The interaction tensor is constructed from P stored patterns ξ_i^μ, where $\mu \in \{1, 2, \ldots, P\}$ indexes patterns via

$$T_{i j_1 j_2 \cdots j_n} = \frac{1}{N^n} \sum_\mu \xi_i^\mu \xi_{j_1}^\mu \xi_{j_2}^\mu \cdots \xi_{j_n}^\mu. \tag{2}$$

The memory load parameter, which appears in the dynamics Eq. (1), is defined as

$$\alpha = \frac{P}{N^n}. \tag{3}$$

This formulation involves $(n+1)$-way interactions among neurons. For $n = 1$, we recover the Hopfield model with pairwise interactions through the matrix

$$T_{ij} = \frac{1}{N} \sum_\mu \xi_i^\mu \xi_j^\mu, \tag{4}$$

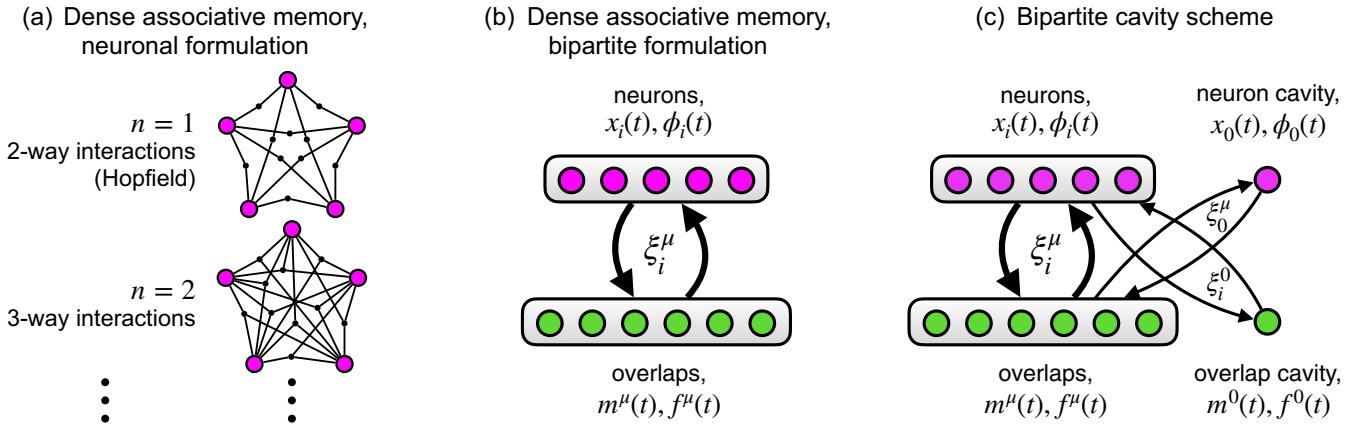

FIG. 1. Schematics of the dense associative memory model. (a) Neuronal formulation with higher-order interactions. Nodes represent neurons, and black dots indicate connections (tensor elements $T_{i j_1 \cdots j_n}$). (b) Equivalent formulation as a bipartite network with neurons $x_i(t)$ and overlaps $m^\mu(t)$ connected in a bipartite manner through stored patterns ξ_i^μ. (c) Schematic of the bipartite cavity scheme used to derive the DMFT.

which can be formed via Hebbian learning with ξ_j^μ and ξ_i^μ representing pre- and postsynaptic activity, respectively.

As $\Delta t \to 0$ while holding the total time $T\Delta t$ fixed, we obtain the continuous-time limit of the dynamics,

$$(1+\partial_t)x_i(t) = \frac{g}{\sqrt{\alpha}} \sum_{j_1, j_2, \cdots, j_n} T_{i j_1 j_2 \cdots j_n} \times \phi_{j_1}(t)\phi_{j_2}(t)\cdots\phi_{j_n}(t) + I_i(t), \tag{5}$$

for which an energy function $\varepsilon[\vec{\phi}]$ can be defined (Sec. III D) [11,12]. Note that these dynamics are not a gradient flow, since the left-hand side specifies $\partial_t x_i(t)$ while the right-hand side is a gradient with respect to ϕ_i. Nevertheless, $\varepsilon[\vec{\phi}]$ serves as a Lyapunov function, decreasing monotonically along trajectories due to the monotonicity of $\phi(\cdot)$ (see Sec. III D for details).

B. Bipartite formulation

This system with higher-order interactions among neurons can be equivalently represented as a bipartite system of neurons and overlaps Fig. 1(b). This reformulation proves advantageous for several reasons. First, it provides a pathway for implementing such models using biological neurons and synapses [19]. Second, the bipartite structure lends itself to the cavity method, which we use to derive the DMFT Fig. 1(c).

We introduce P overlaps $m^\mu(t)$ defined as

$$m^\mu(t) = \frac{1}{N}\sum_i \xi_i^\mu \phi_i(t) + I^\mu(t), \tag{6}$$

where $I^\mu(t)$ are external inputs to the overlaps that serve as source terms in the mean-field analysis. Like the neuronal source terms, they can be set to zero when not needed for this purpose. The overlap $m^\mu(t)$ measures the alignment between the network state at time t, $\phi_i(t)$, and the μth stored pattern, ξ_i^μ. In analogy to the neuronal nonlinearity, we define nonlinearly transformed overlaps as $f^\mu(t) = f(m^\mu(t))$, where $f(m)$ is a monomial:

$$f(m) = m^n, \quad n \geqslant 1. \tag{7}$$

The neuronal dynamics of Eq. (1) can then be written as

$$x_i(t) = (1-\Delta t)x_i(t-1) + \Delta t\left[\frac{g}{\sqrt{\alpha}}\sum_\mu \xi_i^\mu f^\mu(t-1) + I_i(t-1)\right]. \tag{8}$$

In this bipartite representation, neurons and overlaps interact through couplings given by the stored patterns ξ_i^μ. The Hopfield model corresponds to the case where nonlinearity appears only in the neuronal variables [since, in this case, $f(m) = m$]. The dense associative memory model is thus a natural generalization where a pointwise nonlinearity $f(m)$ is also applied to the overlaps.

An interesting extension of these models, which we do not pursue here, is to give the overlaps their own relaxational dynamics rather than the instantaneous equation (6). Moreover, the overlaps can have more complex nonlinearities, including nonlinearities that couple different overlaps to each other, while preserving the existence of an energy function. This reveals connections to self-attention mechanisms [12].

C. Pattern and initial-condition statistics

To analyze the large-N limit, we specify a generative process for the patterns ξ_i^μ, which act as quenched disorder. We adopt the standard assumption of independent and identically distributed pattern components:

$$\xi_i^\mu \overset{\text{iid}}{\sim} P(\xi), \tag{9}$$

where $P(\xi)$ is a probability distribution with zero mean and variance σ_ξ^2.

To study pattern retrieval, we initialize the network with significant overlap with a finite number of patterns of interest, plus random noise. Let $\mu^* \in \{1, 2, \ldots, s\}$ index s patterns of interest, where s is finite and typically small (e.g., $s = 1$). Neurons are initialized as

$$x_i(1) = \sum_{\mu^*} a^{\mu^*}\xi_i^{\mu^*} + z_i, \tag{10}$$

where a^{μ^*} are coefficients determining the initial overlap with pattern μ^*, and z_i represents random noise independent of the patterns:

$$z_i \overset{\text{iid}}{\sim} P(z), \tag{11}$$

with $P(z)$ having zero mean and variance σ_z^2.

D. Condensed patterns and capacity

We now justify the capacity scaling $P = O(N^n)$ through a signal-to-noise analysis. For this scaling to yield interesting dynamics [e.g., phase transitions at $O(1)$ values of α], the signal from the finite number of patterns of interest and the noise from all other patterns encoded in the weights must compete on equal footing.

A key insight underlying the work of Amit, Gutfreund, and Sompolinsky [3,4] is that pattern overlaps have two possible scalings with N, which they referred to as condensed and uncondensed patterns. These correspond to signal and noise, respectively:

$$m^\mu(t) = \begin{cases} O(1) & \text{condensed patterns} \\ O(1/\sqrt{N}) & \text{uncondensed patterns.} \end{cases} \tag{12}$$

The $O(1)$ overlap signifies that the neuronal state is nontrivially aligned with the corresponding condensed pattern. By contrast, the $O(1/\sqrt{N})$ overlap for an uncondensed pattern corresponds to the typical inner product, divided by N, between two random, independent vectors of dimension N. While this scaling is characteristic of independent vectors, uncondensed patterns *do* influence the neuronal state—assuming complete independence between the neuronal state and all patterns would yield incorrect mean-field equations. Condensed and uncondensed patterns are loosely analogous to the rich and lazy regimes of neural-network activity, respectively [20]. The s patterns used for network initialization (Sec. II C) are the condensed patterns, since they start with $O(1)$ overlap and maintain this scaling throughout the dynamics. The remaining patterns stay uncondensed, as patterns initialized with $O(1/\sqrt{N})$ overlap cannot transition to condensed status on $O(1)$ timescales.

Recall that the neuronal input from stored patterns is given by $\frac{g}{\sqrt{\alpha}} \sum_\mu \xi_i^\mu f^\mu(t)$, Eq. (8). There are s condensed patterns and $P - s$ uncondensed patterns, where s is finite, $P = \alpha N^n$, and $N \to \infty$. For condensed patterns, we have $f^\mu(t) = [m^\mu(t)]^n = O(1)$, giving an $O(1)$ contribution to the neuronal input. Each uncondensed pattern contributes much less individually than condensed patterns, but there are many more of them. Since $m^\mu(t) = O(1/\sqrt{N})$ for uncondensed patterns, we have $f^\mu(t) = [m^\mu(t)]^n = O(1/N^{n/2})$. To estimate the total contribution from uncondensed patterns, we treat the quenched disorder ξ_i^μ and dynamic variables $f^\mu(t)$ as independent. While this approximation would yield incorrect mean-field equations, as mentioned above, it suffices for determining the correct scaling behavior. The input from uncondensed patterns has zero mean, and its magnitude scales as

$$\underbrace{\frac{g}{\sqrt{\alpha}}}_{\text{prefactor}} \times \underbrace{\sqrt{P}}_{\text{num. terms in sum}} \times \underbrace{\frac{\sigma_\xi}{N^{n/2}}}_{\text{size of each term}} = g\sigma_\xi, \tag{13}$$

where we have used $P = \alpha N^n$. This shows that the noise contribution from uncondensed patterns is also $O(1)$, confirming that signal and noise compete on equal footing for the chosen scaling $P = \alpha N^n$.

III. DYNAMICAL MEAN-FIELD THEORY (DMFT)

To analyze the transient dynamics of these models in the large-N limit, we develop a DMFT. Unlike traditional equilibrium approaches that focus on fixed points and their stability, DMFT captures the full time evolution of the system, including out-of-equilibrium states where the most interesting memory retrieval properties emerge.

A. Order parameters

The DMFT involves three types of order parameters that characterize macroscopic network activity. The first is the two-time correlation function of neuronal activations,

$$C^\phi(t, t') = \frac{1}{N} \sum_i \phi_i(t)\phi_i(t'). \tag{14}$$

The second is the response function,

$$S^\phi(t, t') = \frac{1}{N} \sum_i \frac{d\phi_i(t)}{dI_i(t')}, \tag{15}$$

which measures how neuronal activations at time t respond to infinitesimal perturbations of the source term $I_i(t')$ at time t'. The third consists of the overlaps with the s condensed patterns used to initialize the dynamics,

$$m^{\mu^*}(t) = \frac{1}{N} \sum_i \xi_i^{\mu^*} \phi_i(t). \tag{16}$$

The DMFT consists of self-consistent equations that determine these order parameters in the limit $N \to \infty$. Finite-size, large-N simulations are expected to match these limiting values up to $O(1/\sqrt{N})$ fluctuations.

B. Approaches

Two main approaches exist for deriving DMFT equations for disordered dynamical systems: path-integral methods and cavity methods. Both approaches address the central challenge that quenched disorder (stored patterns) and dynamic variables (neurons and overlaps) are correlated, making naive disorder averaging incorrect. In the Discussion, we review path integral methods in detail, as they are the basis of most prior work on the DMFT of Hopfield models.

In this paper, we use the cavity method, which provides a more intuitive approach for handling correlations between quenched disorder and dynamic variables. The basic idea is to remove a dynamic variable from the system, creating the titular cavity, then reintroduce it to analyze its effect on the network perturbatively. This approach is particularly well-suited for bipartite systems [21,22] like our neuron-overlap formulation.

The cavity procedure consists of four steps [Fig. 1(c)]:

(1) Begin with an unperturbed system of dynamic variables for a given realization of quenched disorder.

(2) Couple a new "cavity" variable to the existing variables via new random couplings. The cavity variable's introduction perturbs the existing variables.

(3) Write the dynamic equation for the cavity variable, where the input it receives from other variables accounts for how those variables are perturbed in response to the cavity variable's introduction. This perturbation generates a self-coupling term in the resulting single-site dynamics.

(4) Average over the quenched disorder to obtain statistics of the quantities appearing in this single-site picture. The cavity construction allows these averages to be computed because, in the expressions of interest, the relevant quenched disorder—the new random couplings between the cavity variable and the original system—is independent of the dynamic variables. This independence holds because the dynamic variables were defined for the original, unperturbed system, before these new couplings were introduced.

The bipartite structure of the system further simplifies this analysis. When introducing a cavity variable (either a neuron or an overlap), we only need to compute its effect on the opposite type of variables (overlaps or neurons, respectively), since only the opposite type provides direct input to the cavity variable. We perform the cavity analysis twice—once with a neuron cavity and once with an overlap cavity—producing two complementary pictures. The self-consistent equations in each picture depend on statistical averages from the other, creating a closed, mutually referential system that determines the order parameters.

The calculation used here is "zero temperature" in the sense that the dynamic variables follow deterministic evolution given the quenched disorder. Such zero-temperature cavity methods [23] have been applied to static problems, including problems with a bipartite structure [24,25]. For a cavity calculation of the Hopfield equilibrium properties at finite temperature, see Ref. [26].

C. Derivation using the bipartite cavity method

1. Neuron cavity

We first add a "cavity neuron" $x_0(t)$ with activation $\phi_0(t)$ to the system. This neuron connects to all existing overlaps through new random couplings ξ_0^μ. The addition of this neuron perturbs the overlaps of uncondensed patterns by

$$\delta f^\mu(t) = \sum_{t'} \sum_{\nu} \frac{df^\mu(t)}{dI^\nu(t')} \frac{1}{N} \xi_0^\nu \phi_0(t'). \tag{17}$$

The dynamic equation for the cavity neuron, including feedback in response to its own presence, is

$$x_0(t) = (1-\Delta t)x_0(t-1) + \Delta t \left[\underbrace{\frac{g}{\sqrt{\alpha}} \sum_{\mu^*} \xi_0^{\mu^*} f^{\mu^*}(t-1)}_{\text{from condensed patterns}} + \underbrace{\frac{g}{\sqrt{\alpha}} \sum_{\mu} \xi_0^\mu f^\mu(t-1)}_{\substack{=\eta_0(t-1),\\ \text{neuronal cavity field}}} + \sum_{t'} \underbrace{\left[\frac{g}{\sqrt{\alpha}N} \sum_{\mu,\nu} \xi_0^\mu \xi_0^\nu \frac{df^\mu(t-1)}{dI^\nu(t')} \right]}_{\substack{=F_{00}(t-1,t'),\\ \text{neuronal self-coupling kernel}}} \phi_0(t') + I_0(t-1) \right], \tag{18}$$

where we have separated the contributions from condensed patterns μ^* and defined the neuronal cavity field and self-coupling kernel.

The key advantage of the cavity construction is that, as described in step (4) above, it decouples the quenched disorder ξ_0^μ from the dynamic variables $f^\mu(t)$ and $df^\mu(t)/dI^\nu(t')$, allowing us to evaluate disorder-averaged moments of the cavity field and self-coupling kernel. Here and throughout the derivation, $\langle\cdot\rangle$ denotes an average over the quenched disorder, namely, the random patterns ξ_i^μ, ξ_i^0, and ξ_0^μ. By the central limit theorem, the neuronal cavity field $\eta_0(t)$ is Gaussian with statistics

$$\langle \eta_0(t) \rangle = \frac{g}{\sqrt{\alpha}} \sum_{\mu} \underbrace{\langle \xi_0^\mu \rangle}_{=0} \langle f^\mu(t) \rangle = 0, \tag{19}$$

$$\begin{aligned} \langle \eta_0(t)\eta_0(t') \rangle &= \frac{g^2}{\alpha} \sum_{\mu,\nu} \underbrace{\langle \xi_0^\mu \xi_0^\nu \rangle}_{=\delta^{\mu\nu}\sigma_\xi^2} \langle f^\mu(t) f^\nu(t') \rangle \\ &= g^2 \sigma_\xi^2 N^n \langle f^\mu(t) f^\mu(t') \rangle. \end{aligned} \tag{20}$$

Noting that $f^\mu(t) = O(1/N^{n/2})$, the correlation function $\langle \eta_0(t)\eta_0(t') \rangle$ is $O(1)$, meaning that $\eta_0(t)$ itself is $O(1)$, as expected from the scaling arguments in Sec. II D (in which the decoupling achieved here using the cavity construction was instead incorrectly assumed). Meanwhile, the self-coupling kernel $F_{00}(t,t')$ is $O(1)$ and self-averaging with mean

$$\begin{aligned} F_{00}(t,t') &= \frac{g}{\sqrt{\alpha}N} \sum_{\mu,\nu} \underbrace{\langle \xi_0^\mu \xi_0^\nu \rangle}_{=\delta^{\mu\nu}\sigma_\xi^2} \left\langle \frac{df^\mu(t)}{dI^\nu(t')} \right\rangle \\ &= g\sigma_\xi^2 \sqrt{\alpha} N^{n-1} \left\langle \frac{df^\mu(t)}{dI^\mu(t')} \right\rangle. \end{aligned} \tag{21}$$

These averages depend on pattern statistics, which we determine through the complementary overlap-cavity analysis:

$$\langle \eta_0(t)\eta_0(t') \rangle = g^2 \sigma_\xi^2 N^n \langle f^0(t) f^0(t') \rangle, \tag{22}$$

$$F_{00}(t,t') = g\sigma_\xi^2 \sqrt{\alpha} N^{n-1} \left\langle \frac{df^0(t)}{dI^0(t')} \right\rangle. \tag{23}$$

To complete the analysis, we now derive the pattern overlap-cavity equations that will allow us to compute these averages.

2. Overlap cavity

We add an overlap $m^0(t)$, for an uncondensed pattern, connected to all neurons through new random couplings ξ_i^0. The perturbation to neurons due to this cavity overlap is

$$\delta\phi_i(t) = \sum_{t'}\sum_j \frac{d\phi_i(t)}{dI_j(t')}\frac{g}{\sqrt{\alpha}}\xi_j^0 f^0(t'). \tag{24}$$

The dynamic equation for the cavity overlap is then

$$m^0(t) = \underbrace{\frac{1}{N}\sum_i \xi_i^0\phi_i(t)}_{=\eta^0(t),\ \text{overlap cavity field}} + \sum_{t'}\underbrace{\left[\frac{g}{\sqrt{\alpha}N}\sum_{i,j}\xi_i^0\xi_j^0\frac{d\phi_i(t)}{dI_j(t')}\right]}_{=F^{00}(t,t'),\ \text{overlap self-coupling kernel}} f^0(t') + I^0(t), \tag{25}$$

where we have defined the overlap cavity field and self-coupling kernel. As with the neuronal cavity, we use the independence of quenched disorder and dynamic variables to evaluate disorder averages. The cavity field $\eta^0(t)$ is Gaussian with statistics

$$\langle\eta^0(t)\rangle = \frac{1}{N}\sum_i \underbrace{\langle\xi_i^0\rangle}_{=0}\langle\phi_i(t)\rangle = 0, \tag{26}$$

$$\langle\eta^0(t)\eta^0(t')\rangle = \frac{1}{N^2}\sum_{i,j}\underbrace{\langle\xi_i^0\xi_j^0\rangle}_{=\delta_{ij}\sigma_\xi^2}\langle\phi_i(t)\phi_j(t')\rangle = \frac{\sigma_\xi^2}{N}C^\phi(t,t'). \tag{27}$$

The correlation function $\langle\eta^0(t)\eta^0(t')\rangle$ is $O(1/N)$, implying that $\eta^0(t) = O(1/\sqrt{N})$, consistent with the overlap itself being $O(1/\sqrt{N})$ as expected for an uncondensed pattern. Meanwhile, the self-coupling kernel is $O(1)$ and self-averaging with mean

$$F^{00}(t,t') = \frac{g}{\sqrt{\alpha}N}\sum_{i,j}\underbrace{\langle\xi_i^0\xi_j^0\rangle}_{=\delta_{ij}\sigma_\xi^2}\left\langle\frac{d\phi_i(t)}{dI_j(t')}\right\rangle = \frac{g\sigma_\xi^2}{\sqrt{\alpha}}S^\phi(t,t'). \tag{28}$$

Thus, the overlap cavity picture's cavity field and self-coupling kernel depend on the neuronal order parameters $C^\phi(t,t')$ and $S^\phi(t,t')$, which can be determined within the neuronal cavity picture. This creates mutually referential cavity pictures that together determine the order parameters.

3. Evaluating correlation and response functions

Since we aim to determine the neuronal order parameters $C^\phi(t,t')$, $S^\phi(t,t')$, and $m^{\mu^*}(t)$, it is useful to close the mean-field equations in neuronal quantities. To do this, we must evaluate the neuronal cavity-field correlation Eq. (22) and self-coupling kernel Eq. (23). To deal with temporal indices, it is helpful to use the following vector and matrix notation:

(1) For a time-dependent scalar quantity $q(t)$ with $t \in \{1,2,\ldots,T\}$, we define the corresponding T-dimensional vector $\boldsymbol{q}$ with components $[\boldsymbol{q}]_t = q(t)$.

(2) For a two-time function $M(t,t')$ with $t,t' \in \{1,2,\ldots,T\}$, we define the corresponding $T\times T$ matrix $\boldsymbol{M}$ with elements $[\boldsymbol{M}]_{t,t'} = M(t,t')$.

(3) For a two-time derivative $d\psi(t)/dI(t')$, we define the corresponding $T\times T$ matrix $d\boldsymbol{\psi}/d\boldsymbol{I}^T$ with elements $[d\boldsymbol{\psi}/d\boldsymbol{I}^T]_{t,t'} = d\psi(t)/dI(t')$.

In this notation, Eqs. (22) and (23) for the neuronal cavity-field correlation and self-coupling kernel, respectively, are

$$\langle\boldsymbol{\eta}_0\boldsymbol{\eta}_0^T\rangle = g^2\sigma_\xi^2 N^n\langle f(\boldsymbol{m}^0)f(\boldsymbol{m}^0)^T\rangle, \tag{29}$$

$$\boldsymbol{F}_{00} = g\sigma_\xi^2\sqrt{\alpha}N^{n-1}\left\langle\frac{df(\boldsymbol{m}^0)}{d(\boldsymbol{I}^0)^T}\right\rangle, \tag{30}$$

where $f(\boldsymbol{m}^0)$ applies the nonlinearity elementwise to the vector $\boldsymbol{m}^0$. We need to evaluate these expressions to leading order, namely, $O(1)$. From the overlap cavity picture, $\boldsymbol{m}^0$ obeys, in matrix notation,

$$\boldsymbol{m}^0 = \boldsymbol{\eta}^0 + \frac{g\sigma_\xi^2}{\sqrt{\alpha}}\boldsymbol{S}^\phi f(\boldsymbol{m}^0) + \boldsymbol{I}^0, \tag{31}$$

$$\text{where}\quad \langle\boldsymbol{\eta}^0\rangle = 0, \quad \langle\boldsymbol{\eta}^0(\boldsymbol{\eta}^0)^T\rangle = \frac{\sigma_\xi^2}{N}\boldsymbol{C}^\phi. \tag{32}$$

Here, $\boldsymbol{\eta}^0$ is Gaussian and $\boldsymbol{m}^0$ is determined by solving the nonlinear equation (31).

At this point, the analysis diverges between the Hopfield ($n=1$) and higher-order models ($n>1$). In the $n=1$ case, the overlap equation (31) is linear, which simplifies the calculations. For $n>1$, Eq. (31) is nonlinear, but the nonlinear self-interaction is smaller than $\boldsymbol{\eta}^0$ by a factor of $1/N^{(n-1)/2}$, allowing for a perturbative treatment.

Before handling each case, we derive a general expression for the response-function term $\langle d\boldsymbol{m}^0/d(\boldsymbol{I}^0)^T\rangle$ that applies to both cases. Differentiating both sides of Eq. (31) with respect to $\boldsymbol{I}^0$ and solving for $d\boldsymbol{m}^0/d(\boldsymbol{I}^0)^T$ gives

$$\frac{d\boldsymbol{m}^0}{d(\boldsymbol{I}^0)^T} = \left(\boldsymbol{\mathcal{I}} - \frac{g\sigma_\xi^2}{\sqrt{\alpha}}\boldsymbol{S}^\phi\mathcal{D}[f'(\boldsymbol{m}^0)]\right)^{-1}, \tag{33}$$

where $\mathcal{D}[\cdot]$ denotes a diagonal matrix with the argument vector on the diagonal when applied to a vector, or zeros out the off-diagonal elements when applied to a matrix; $\boldsymbol{\mathcal{I}}$ is the identity matrix; and we set $\boldsymbol{I}^0 = \boldsymbol{0}$. This allows us to express Eq. (30) as

$$\boldsymbol{F}_{00} = g\sigma_\xi^2\sqrt{\alpha}N^{n-1} \times\left\langle\mathcal{D}[f'(\boldsymbol{m}^0)]\left(\boldsymbol{\mathcal{I}} - \frac{g\sigma_\xi^2}{\sqrt{\alpha}}\boldsymbol{S}^\phi\mathcal{D}[f'(\boldsymbol{m}^0)]\right)^{-1}\right\rangle. \tag{34}$$

Hopfield model ($n = 1$). For the Hopfield model, due to the linearity of the overlap dynamics, we have

$$\boldsymbol{m}^0 = \frac{d\boldsymbol{m}^0}{d(\boldsymbol{I}^0)^T}\boldsymbol{\eta}^0, \tag{35}$$

$$\text{where} \quad \frac{d\boldsymbol{m}^0}{d(\boldsymbol{I}^0)^T} = \left(\mathcal{I} - \frac{g\sigma_\xi^2}{\sqrt{\alpha}}\boldsymbol{S}^\phi\right)^{-1}. \tag{36}$$

Thus, Eq. (29) becomes

$$\begin{aligned}\left\langle \boldsymbol{\eta}_0\boldsymbol{\eta}_0^T\right\rangle &= g^2\sigma_\xi^2 N\langle \boldsymbol{m}^0(\boldsymbol{m}^0)^T\rangle \\ &= g^2\sigma_\xi^2 N\left\langle \frac{d\boldsymbol{m}^0}{d(\boldsymbol{I}^0)^T}\boldsymbol{\eta}^0(\boldsymbol{\eta}^0)^T\left(\frac{d\boldsymbol{m}^0}{d(\boldsymbol{I}^0)^T}\right)^T\right\rangle \\ &= g^2\sigma_\xi^4\left(\mathcal{I} - \frac{g\sigma_\xi^2}{\sqrt{\alpha}}\boldsymbol{S}^\phi\right)^{-1}\boldsymbol{C}^\phi\left(\mathcal{I} - \frac{g\sigma_\xi^2}{\sqrt{\alpha}}\boldsymbol{S}^\phi\right)^{-T}. \end{aligned} \tag{37}$$

Meanwhile, since $f(m) = m$, Eq. (34) simplifies to

$$\boldsymbol{F}_{00} = g\sigma_\xi^2\sqrt{\alpha}\left(\mathcal{I} - \frac{g\sigma_\xi^2}{\sqrt{\alpha}}\boldsymbol{S}^\phi\right)^{-1}. \tag{38}$$

Higher-order models ($n > 1$). For $n > 1$, the pattern equation (31) is nonlinear, but, as mentioned above, the nonlinear self-coupling term $f(\boldsymbol{m}^0)$ is smaller than the Gaussian input $\boldsymbol{\eta}^0$ by a factor of $1/N^{(n-1)/2}$, permitting a perturbative treatment. For Eq. (29), to leading order, we can replace $\boldsymbol{m}^0$ with $\boldsymbol{\eta}^0$:

$$\begin{aligned}\left\langle \boldsymbol{\eta}_0\boldsymbol{\eta}_0^T\right\rangle &= g^2\sigma_\xi^2 N^n\langle f(\boldsymbol{\eta}^0)f(\boldsymbol{\eta}^0)^T\rangle \\ &= g^2\sigma_\xi^{2(n+1)}\boldsymbol{P}_{n,n}, \end{aligned} \tag{39}$$

where we define the matrices

$$\boldsymbol{P}_{n,n'} = \langle \boldsymbol{u}^n(\boldsymbol{u}^{n'})^T\rangle_{\boldsymbol{u}\sim\mathcal{N}(0,\boldsymbol{C}^\phi)}, \tag{40}$$

with powers applied elementwise.

For the response function, noting that $f'(\boldsymbol{m}^0) = O(1/N^{(n-1)/2})$, we expand the matrix inverse in Eq. (34):

$$\begin{aligned}\boldsymbol{F}_{00} &= g\sigma_\xi^2\sqrt{\alpha}N^{n-1}\left\langle \mathcal{D}[f'(\boldsymbol{m}^0)]\left(\mathcal{I} + \frac{g\sigma_\xi^2}{\sqrt{\alpha}}\boldsymbol{S}^\phi\mathcal{D}[f'(\boldsymbol{m}^0)] + \cdots\right)\right\rangle \\ &= g\sigma_\xi^2\sqrt{\alpha}N^{n-1}\langle\mathcal{D}[f'(\boldsymbol{m}^0)]\rangle + g^2\sigma_\xi^4\boldsymbol{S}^\phi\circ\langle N^{n-1}f'(\boldsymbol{m}^0)f'(\boldsymbol{m}^0)^T\rangle + \cdots, \end{aligned} \tag{41}$$

where $\circ$ denotes the Hadamard product.

For odd n, $f'(m) = nm^{n-1}$ is an even function, so $\langle f'(\boldsymbol{m}^0)\rangle \neq 0$ and is $O(1/N^{(n-1)/2})$. The first term in the expansion Eq. (41) then scales as $O(N^{(n-1)/2})$, which diverges. This divergence could presumably be removed through exclusion of self-interactions in Eq. (1). We leave this to future work and restrict to even n, for which f' is odd, and the problematic term vanishes. The different scaling behaviors of the neuronal input for even and odd n can be illustrated through a simple numerical experiment (Appendix D, Fig. 7).

For even n, we use Eq. (31) to iteratively express $f'(\boldsymbol{m}^0)$ in terms of $\boldsymbol{\eta}^0$:

$$f'(\boldsymbol{m}^0) = f'(\boldsymbol{\eta}^0) + \frac{g\sigma_\xi^2}{\sqrt{\alpha}}\mathcal{D}[f''(\boldsymbol{\eta}^0)]\boldsymbol{S}^\phi f(\boldsymbol{\eta}^0) + \cdots, \tag{42}$$

whose kth term is $O(1/N^{k(n-1)/2})$. Substitution into the first term of the $\boldsymbol{F}_{00}$ expansion Eq. (41) gives

$$g\sigma_\xi^2\sqrt{\alpha}N^{n-1}\left\langle\mathcal{D}[f'(\boldsymbol{m}^0)]\right\rangle = g\sigma_\xi^2\sqrt{\alpha}N^{n-1}\left\langle\mathcal{D}[f'(\boldsymbol{\eta}^0)]\right\rangle + g^2\sigma_\xi^4 N^{n-1}\mathcal{D}[\boldsymbol{S}^\phi\left\langle f(\boldsymbol{\eta}^0)f''(\boldsymbol{\eta}^0)^T\right\rangle] + \cdots. \tag{43}$$

Since n is even, $f'(m)$ is an odd function, so $\langle f'(\boldsymbol{\eta}^0)\rangle = 0$. For the second term in the expansion Eq. (41), we need $\langle N^{n-1}f'(\boldsymbol{m}^0)f'(\boldsymbol{m}^0)^T\rangle$. To leading order, we can replace $f'(\boldsymbol{m}^0)$ with $f'(\boldsymbol{\eta}^0)$. Combining these results gives

$$\boldsymbol{F}_{00} = g^2\sigma_\xi^4[\mathcal{D}[\boldsymbol{S}^\phi\langle N^{n-1}f(\boldsymbol{\eta}^0)f''(\boldsymbol{\eta}^0)^T\rangle] + \boldsymbol{S}^\phi\circ\langle N^{n-1}f'(\boldsymbol{\eta}^0)f'(\boldsymbol{\eta}^0)^T\rangle], \tag{44}$$

where $\mathcal{D}[\cdot]$ now extracts the diagonal elements. This can be written as

$$\boldsymbol{F}_{00} = g^2\sigma_\xi^{2(n+1)}[n(n-1)\mathcal{D}[\boldsymbol{S}^\phi\boldsymbol{P}_{n,n-2}] + n^2\boldsymbol{S}^\phi\circ\boldsymbol{P}_{n-1,n-1}]. \tag{45}$$

4. Final self-consistent system

In the single-site picture, $x_0(t)$ evolves from an initial condition, driven by a cavity field and self-coupling. The statistics of this single-site process determine the self-coupling kernel and cavity-field correlation function:

$$x_0(1) = \sum_{\mu^*=1} a^{\mu^*}\xi_0^{\mu^*} + z_0, \tag{46}$$

$$x_0(t) = (1-\Delta t)x_0(t-1) + \Delta t\left[\frac{g}{\sqrt{\alpha}}\sum_{\mu^*=1}\xi_0^{\mu^*}[m^{\mu^*}(t-1)]^n + \eta_0(t-1) + \sum_{t'=1}^{t-1}F_{00}(t-1,t')\phi_0(t')\right], \tag{47}$$

$$\boldsymbol{C}^{\eta_0} = \begin{cases} g^2\sigma_\xi^4\left(\mathcal{I} - \frac{g\sigma_\xi^2}{\sqrt{\alpha}}\boldsymbol{S}^\phi\right)^{-1}\boldsymbol{C}^\phi\left(\mathcal{I} - \frac{g\sigma_\xi^2}{\sqrt{\alpha}}\boldsymbol{S}^\phi\right)^{-T} & n = 1 \\ g^2\sigma_\xi^{2(n+1)}\boldsymbol{P}_{n,n} & n > 1, \text{ even}, \end{cases} \tag{48}$$

$$\boldsymbol{F}_{00} = \begin{cases} g\sigma_\xi^2\sqrt{\alpha}\left(\mathcal{I} - \frac{g\sigma_\xi^2}{\sqrt{\alpha}}\boldsymbol{S}^\phi\right)^{-1} & n = 1 \\ g^2\sigma_\xi^{2(n+1)}\{n(n-1)\mathcal{D}[\boldsymbol{S}^\phi\boldsymbol{P}_{n,n-2}] + n^2\boldsymbol{S}^\phi \circ \boldsymbol{P}_{n-1,n-1}\} & n > 1, \text{ even}. \end{cases} \tag{49}$$

The DMFT is closed by the following self-consistency conditions, with the single-site average $\langle\cdot\rangle_{\text{single-site}}$ defined in Eq. (53):

$$C^\phi(t,t') = \left\langle \phi_0(t)\phi_0(t')\right\rangle_{\text{single-site}}, \tag{50}$$

$$S^\phi(t,t') = \left\langle \frac{d\phi_0(t)}{dI_0(t')}\right\rangle_{\text{single-site}}, \tag{51}$$

$$m^{\mu^*}(t) = \left\langle \xi_0^{\mu^*}\phi_0(t)\right\rangle_{\text{single-site}}, \tag{52}$$

where the single-site average $\langle\cdot\rangle_{\text{single-site}}$ denotes averaging over the Gaussian noise realization, condensed pattern components, and initialization noise:

$$\langle\cdots\rangle_{\text{single-site}} = \langle\cdots\rangle_{\substack{\eta_0\sim N(\mathbf{0},\mathbf{C}^{\eta_0}) \\ \xi_0^{\mu^*}\overset{\text{iid}}{\sim}P(\xi)\text{ for }\mu^*=1,\ldots,s \\ z_0\sim P(z)}}. \tag{53}$$

These equations form a closed system that can be solved numerically to obtain the dynamical behavior of the model in the $N\to\infty$ limit.

D. Energy function

The models we study possess energy functions that govern their dynamics [11,12]. For a configuration of neuronal activations $\vec{\phi} = \{\phi_i\}_{i=1}^N$, the $O(1)$ energy is

$$\varepsilon[\vec{\phi}] = -\frac{g}{(n+1)\sqrt{\alpha}}\sum_\mu (m^\mu)^{n+1} + \frac{1}{N}\sum_i \mathcal{F}(\phi_i), \tag{54}$$

where $\mathcal{F}(\phi)$ satisfies $\mathcal{F}'(\phi) = \phi^{-1}(\phi)$ (Appendix A). The continuous-time limit of the dynamics obeys

$$\partial_t x_i(t) = -N\partial_{\phi_i}\varepsilon[\vec{\phi}(t)]. \tag{55}$$

Note that these are not gradient dynamics because the left-hand side specifies the time derivative of x_i while the right-hand side is a gradient with respect to ϕ_i. Nevertheless, $\varepsilon[\vec{\phi}]$ acts as a Lyapunov function since the energy decreases monotonically:

$$\begin{aligned}\partial_t\varepsilon[\vec{\phi}] &= \sum_i \partial_{\phi_i}\varepsilon[\vec{\phi}]\phi_i'(t)\partial_t x_i(t) \\ &= -\frac{1}{N}\sum_i \phi_i'(t)[\partial_t x_i(t)]^2 \leqslant 0,\end{aligned} \tag{56}$$

where the inequality follows because the nonlinearity is monotonic, $\phi'(x) > 0$. In the DMFT framework, we can express the energy in terms of order parameters as

$$\begin{aligned}\varepsilon(t) = &-\begin{Bmatrix} \frac{\sqrt{\alpha}}{2g}\langle\eta_0(t)\eta_0(t)\rangle_{\text{single-site}} & n = 1 \\ \langle\eta_0(t)\phi_0(t)\rangle_{\text{single-site}} & n > 1, \text{even} \end{Bmatrix} \\ &- \frac{g}{(n+1)\sqrt{\alpha}}\sum_{\mu^*=1}\left[m^{\mu^*}(t)\right]^{n+1} + \langle\mathcal{F}(\phi_0(t))\rangle_{\text{single-site}}.\end{aligned} \tag{57}$$

We found this expression to be the most numerically stable of several equivalent formulations.

E. Numerical solution of the DMFT

We solve the self-consistent DMFT equations using an iterative procedure that samples trajectories and updates order parameters. A very closely related approach for an ecological system is detailed in Ref. [17]. The steps are as follows:

(1) Initialize order parameters $\boldsymbol{C}^\phi$, $\boldsymbol{S}^\phi$, and $\boldsymbol{m}^{\mu^*}$ for $\mu^* \in \{1, 2, \ldots, s\}$.

(2) Sample M noise trajectories $\boldsymbol{\eta}_m$, $m \in \{1, 2, \ldots, M\}$, through Cholesky decomposition of $\boldsymbol{C}^{\eta_0}$.

(3) Sample M sets of s condensed patterns $\xi_m^{\mu^*}$.

(4) Forward integrate the M trajectories using the single-site dynamics to get $\boldsymbol{x}_m$ and $\boldsymbol{\phi}_m = \phi(\boldsymbol{x}_m)$, yielding updated correlation function $\boldsymbol{C}^\phi$ and overlaps $\boldsymbol{m}^{\mu^*}$ in the straightforward way [Eqs. (58) and (59) below].

(5) Compute an updated response function $\boldsymbol{S}^\phi$ (described in Sec. III E 1 below).

(6) Update order parameters with memory factor $\gamma \in [0, 1]$:

$$\boldsymbol{C}^\phi_{\text{new}} = (1-\gamma)\boldsymbol{C}^\phi_{\text{old}} + \gamma\frac{1}{M}\sum_m \boldsymbol{\phi}_m\boldsymbol{\phi}_m^T, \tag{58}$$

$$\boldsymbol{m}^{\mu^*}_{\text{new}} = (1-\gamma)\boldsymbol{m}^{\mu^*}_{\text{old}} + \gamma\frac{1}{M}\sum_m \xi_m^{\mu^*}\boldsymbol{\phi}_m, \tag{59}$$

$$\boldsymbol{S}^\phi_{\text{new}} = (1-\gamma)\boldsymbol{S}^\phi_{\text{old}} + \gamma\boldsymbol{S}^\phi_{\text{tot}}. \tag{60}$$

(7) Repeat steps (2)–(6) until convergence of order parameters.

1. Response function ($\boldsymbol{S}^\phi_{\text{tot}}$) computation

For each trajectory m, the response function is computed through forward integration in t for each fixed s. The

computation begins at $t = s$ and proceeds forward in time:

$$S_m^x(t,s) = (1-\Delta t)S_m^x(t-1,s) + \Delta t\left[\sum_{t'=s}^{t-1} F_{00}(t-1,t')\,\phi_m'(t')\,S_m^x(t',s) + \delta_{t-1,s}\right], \tag{61}$$

subject to initial conditions

$$S_m^x(s,s) = 0. \tag{62}$$

The equation is applied sequentially for $t = s+1, s+2, \ldots, T$ to build up the full response function. For each trajectory, the activation response function is

$$S_m^\phi(t,s) = \phi_m'(t)\,S_m^x(t,s). \tag{63}$$

The final response function is obtained by averaging over all trajectories:

$$S_{\text{tot}}^\phi(t,s) = \frac{1}{M}\sum_m S_m^\phi(t,s). \tag{64}$$

We implement the numerical solution on a GPU using PyTorch, enabling efficient computation with large sample sizes M. The primary computational loops are the T time steps for trajectory evolution and the $T(T-1)/2$ time steps for response function integration. The $O(T^2)$ response function computation dominates the runtime in practice.

IV. RESULTS

A. Simulation setup

We restrict our analysis to interaction orders $n = 1$, 2, and 4. Validation with finite-size simulations requires large N to reduce fluctuations that scale as $O(1/\sqrt{N})$, but the scaling $P = O(N^n)$ makes the number of patterns prohibitively large for $n > 4$. We exclude $n = 3$ due to the divergent behavior identified in Sec. III.

All simulations and DMFT solutions focus on the retrieval of a single condensed pattern, so we drop the μ^* superscript and denote the overlap as $m(t)$. We initialize the system according to Eq. (10) with $a = \bar{a}g$, where $\bar{a} \in [0, 1]$ controls the initial alignment with the pattern of interest. The initialization noise level is set to $\sigma_z = \sqrt{g^2 - a^2}$ [see Eq. (11)] so that the variance of the initial condition remains constant as we vary $\bar{a}$. By sweeping $\bar{a}$ from 0 to 1, we explore a range of initial overlaps with the stored pattern.

For simulations, we use system sizes $N = 20\,000$, 2000, and 200 for $n = 1$, 2, and 4, respectively, with time steps $\Delta t = 0.25$ for $n = 1, 2$ and $\Delta t = 0.05$ for $n = 4$ to ensure numerical stability. We examine the effects of g and Δt in the Hopfield model in Appendix C.

B. Equilibrium analysis

We first confirm that our model exhibits the equilibrium blackout catastrophe—the discontinuous phase transition where high-overlap solutions vanish above a critical capacity. We derive a mean-field theory for fixed-point solutions by removing the time dependencies from our DMFT equations. For the gain parameter $g = 1.5$ used throughout our analyses, we obtain critical capacities $\alpha_c \approx 0.13$, 0.080, and 0.0011 for $n = 1$, 2, and 4, respectively. Details of this fixed-point analysis are provided in Appendix B.

C. Retrieval dynamics and key phenomena

The DMFT solutions show two types of retrieval behaviors.

Stable retrieval occurs when the overlap converges to a value independent of local variations in initial conditions. The convergence is rapid, and the asymptotic overlap is close to unity, as expected from prior equilibrium analyses [3,4,18].

Transient retrieval occurs when the overlap initially increases but then decreases, failing to reach a stable state. This occurs either when networks exceed capacity or when they are initialized outside basins of attraction. The initial increase is fast, while the eventual decay is much slower, and the late-time overlap value appears to remain nonzero, consistent with the "remnant overlap" found in simulations of binary-spin Hopfield models [4]. We cannot determine the exact asymptotic values of this remnant overlap because the $O(T^2)$ time complexity of our numerical solver limits the accessible time horizon.

Unlike in networks of binary spins, the time-dependent variance $C^\phi(t,t)$ is nontrivial, and changes in the overlap $m(t)$ reflect both changes in alignment with the pattern and changes in the variance of the neuronal state. We therefore focus on the normalized overlap,

$$\bar{m}(t) = \frac{m(t)}{\sigma_\xi\sqrt{C^\phi(t,t)}}, \tag{65}$$

which lies in $[-1, 1]$. We define $\bar{m}_{\text{init}} = \bar{m}(1)$ as the initial normalized overlap.

D. Validation and main results

We first verify that DMFT solutions match finite-size simulations. Figure 2 demonstrates excellent agreement between theory and simulations for the overlap $m(t)$, equal-time correlation function $C^\phi(t,t)$, and normalized overlap $\bar{m}(t)$.

The normalized overlap exhibits two regimes:

Below capacity ($\alpha < \alpha_c$). When $\bar{m}_{\text{init}}$ is sufficiently large, $\bar{m}(t)$ converges to a stable retrieval state with rapid convergence and final values close to unity. For smaller $\bar{m}_{\text{init}}$, $\bar{m}(t)$ transiently increases and then decays slowly. Thus, even when initialized outside the basin of attraction, patterns can still be transiently recalled.

Above capacity ($\alpha > \alpha_c$). No initial overlap $\bar{m}_{\text{init}}$ is sufficiently large to elicit stable retrieval. Instead, we observe transient retrieval for all $\bar{m}_{\text{init}}$. However, contrary to the equilibrium picture predicting abrupt breakdown, maximum normalized overlaps achieved can be quite high. Since the transient increase is fast and subsequent decay much slower, this behavior can resemble stable retrieval from the below-capacity regime, with the key difference being eventual slow decay rather than convergence to a stable state.

E. Transient-recovery curves

To quantify memory retrieval performance beyond stable attractors, we introduce transient-recovery curves. These curves characterize a network's ability to recall stored patterns

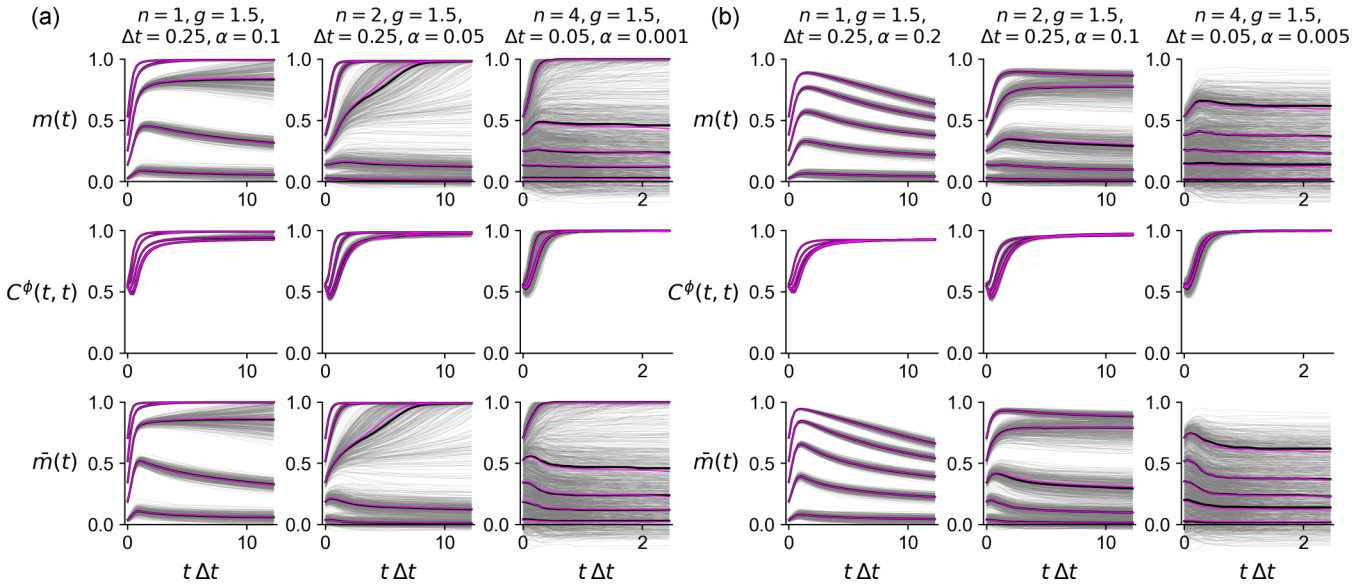

FIG. 2. Dynamical evolution of order parameters for Hopfield ($n = 1$) and dense associative memory models ($n = 2, 4$). (a) Below-capacity dynamics with $\alpha = 0.10, 0.05, 0.001$ for $n = 1, 2, 4$, respectively. (b) Above-capacity dynamics with $\alpha = 0.20, 0.10, 0.005$ for $n = 1, 2, 4$, respectively. In panels (a) and (b), we show (top) raw overlap $m(t)$, (middle) equal-time correlation $C^{\phi}(t, t)$, and (bottom) normalized overlap $\bar{m}(t) = m(t)/[\sigma_{\xi}\sqrt{C^{\phi}(t, t)}]$. Gray traces show individual finite-size simulations with $N = 20\,000, 2000, 200$ for $n = 1, 2, 4$, respectively; black lines show simulation medians; and magenta lines show DMFT predictions.

by plotting the maximum normalized overlap achieved during the entire dynamical evolution,

$$\bar{m}_{\max} = \max_{t} \bar{m}(t), \tag{66}$$

as a function of the initial normalized overlap $\bar{m}_{\text{init}}$. Each curve captures the best retrieval performance accessible through transient dynamics, regardless of whether this optimal retrieval occurs at a stable fixed point or during a transient. Figure 3 shows these transient-recovery curves for the Hopfield model and dense associative memory models with $n = 2$ and $n = 4$, where each curve corresponds to a different memory load α.

All curves lie above the diagonal since $\bar{m}_{\max} \geqslant \bar{m}_{\text{init}}$. When α is sufficiently small and $\bar{m}_{\text{init}}$ is sufficiently large, the network enters a stable retrieval state. This is reflected by the curve becoming flat. That is, $\bar{m}_{\max}$ becomes insensitive to local changes in $\bar{m}_{\text{init}}$. Below capacity but outside the basin of attraction, $\bar{m}_{\max}$ increases smoothly with $\bar{m}_{\text{init}}$ until entering the basin, reflecting transient retrieval. Above the critical threshold, the entire curve shows smooth increases due to transient retrieval.

The transient-recovery curves reveal that going above capacity results in changes to retrieval performance that are far more graceful than equilibrium analyses might suggest. As α increases from below to above critical capacity, the curves change smoothly rather than exhibiting abrupt discontinuities (of course, the presence or absence of a flat plateau represents a qualitative difference between the two regimes). Thus, the blackout catastrophe is considerably less catastrophic when

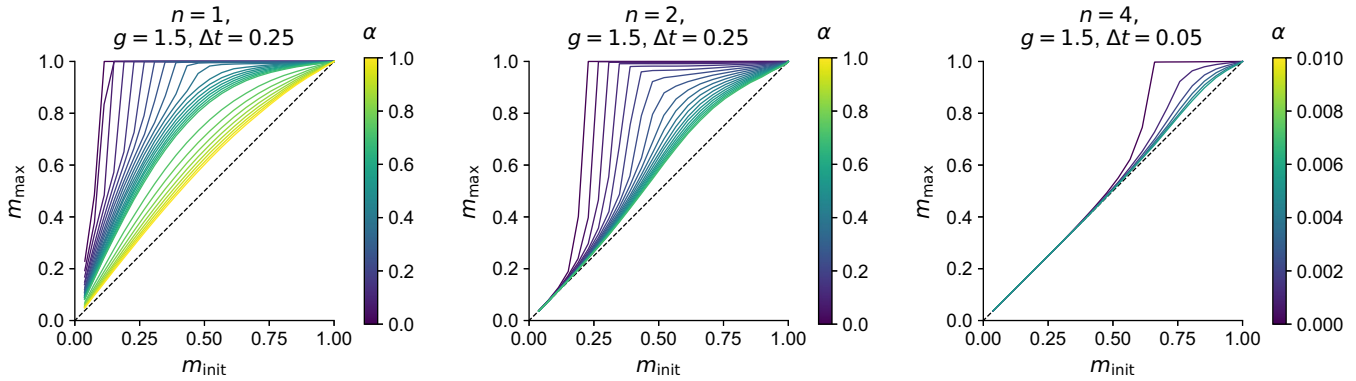

FIG. 3. Transient-recovery curves for Hopfield model ($n = 1$) and dense associative memory models ($n = 2, 4$). Each curve plots the maximum normalized overlap $\bar{m}_{\max}$ achieved during dynamical evolution versus the initial normalized overlap $\bar{m}_{\text{init}}$. Different curves within each panel correspond to different memory loads $\alpha = P/N^n$. The diagonal line is the trivial lower bound where maximum overlap equals initial overlap.

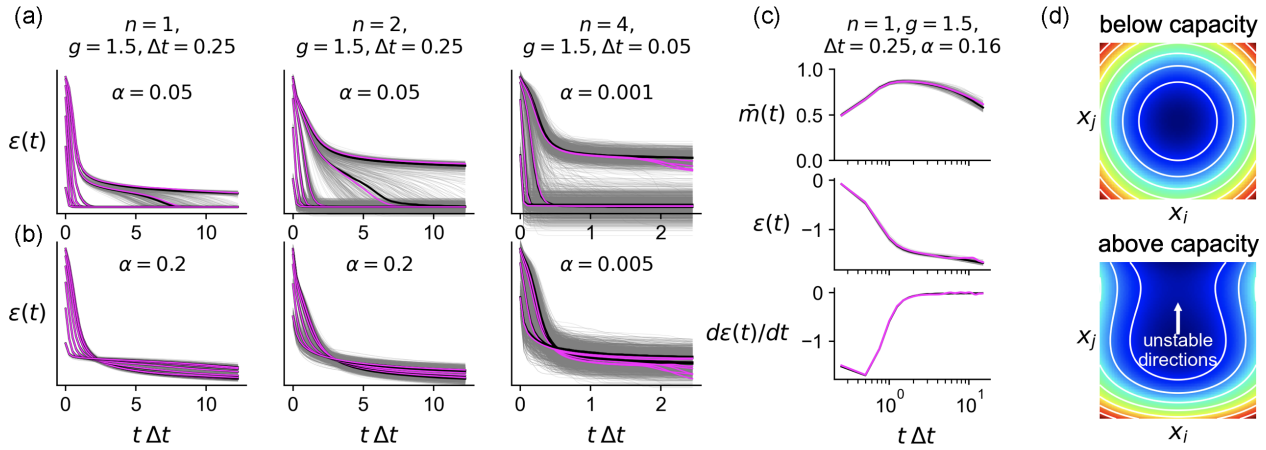

FIG. 4. Energy dynamics for different initial overlaps ($g = 1.5$). (a) Below-capacity dynamics with $\alpha < \alpha_c$. (b) Above-capacity dynamics with $\alpha > \alpha_c$. In panels (a) and (b), columns show $n = 1, 2, 4$ from left to right. Different curves in each panel correspond to different initial normalized overlaps $\bar{m}_{\text{init}}$. Gray traces show individual finite-size simulations; black lines show simulation medians; and magenta lines show DMFT predictions. (c) Example showing correspondence between energy decay and overlap evolution for $n = 1$, $\alpha = 0.16$. (top) normalized overlap $\bar{m}(t)$; (middle) energy $\varepsilon(t)$; (bottom) energy derivative $d\varepsilon(t)/dt$. Horizontal axis shows time on a log scale. The fast rise and slow decay of the normalized overlap correspond to fast decay and slow decay of the energy, consistent with the system navigating shallow energy landscape features near stored patterns that eventually drive it away from the memory. (d) Schematic of the energy landscape structure near a stored pattern in the below-capacity regime, where a stable basin exists (top), and in the above-capacity regime, where a slow, unstable region persists despite the elimination of the stable basin (bottom).

viewed through transient dynamics. Networks retain substantial memory function even beyond their critical capacities, provided one accepts transient rather than persistent recall.

While comparing individual curves at fixed α across different n is not meaningful since α_c varies strongly with n, the families of curves generated by varying α can be compared. Comparing these families across interaction orders reveals important differences in transient retrieval characteristics. The Hopfield model exhibits the most graceful degradation, with curves that maintain roughly symmetric shape around the diagonal as α increases. In contrast, higher-order models show increasingly asymmetric behavior, with transient recovery effects becoming most pronounced for large $\bar{m}_{\text{init}}$. This asymmetry is most extreme for $n = 4$, where large initial overlaps are required to obtain substantial transient retrieval. Conversely, the Hopfield model shows significant transient recovery even for modest initial overlaps.

F. Energy dynamics

We now show that transient retrieval is caused by slow regions in the energy landscape near memories where, below capacity, stable fixed points previously existed.

Figures 4(a) and 4(b) show energy versus time for different initial overlaps, with columns representing $n = 1, 2, 4$ and Figs. 4(a) and 4(b) showing below- and above-capacity regimes, respectively. When entering stable retrieval states (below capacity with sufficient initial overlap), the energy quickly drops to a fixed value independent of the specific initial overlap. When displaying transient retrieval (above capacity or with small initial overlap), the energy exhibits two distinct timescales of decay: fast initial decay followed by slow decay. This slow decay corresponds to the system exploring regions of small gradient (that is, slow regions) of the energy landscape.

The correspondence between energy and overlap dynamics is illustrated in Fig. 4(c), which shows three complementary views: normalized overlap $\bar{m}(t)$ (top), energy $\varepsilon(t)$ (middle), and energy time derivative $d\varepsilon(t)/dt$ (bottom), all shown on a logarithmic timescale. The fast energy decay corresponds to the fast rise of $\bar{m}(t)$, while the slow energy decay corresponds to the slow decay of $\bar{m}(t)$. This demonstrates that transient retrieval occurs when the system becomes trapped in slowly varying regions of the energy landscape—residual features of the stable fixed points that existed below capacity. We schematize this landscape structure in Fig. 4(d).

G. Optimal readout time

In networks storing memories as stable fixed points, memories can be read out at any sufficiently late time after convergence. However, when memories are retrieved via transient dynamics in slow regions of the energy landscape, there exists an optimal readout time that maximizes retrieval performance.

Figure 5 plots the optimal readout time,

$$t_{\text{opt}} = \arg\max_t \bar{m}(t), \tag{67}$$

from the DMFT as a function of $\bar{m}_{\text{init}}$. Each row shows increasing values of α, with columns showing $n = 1, 2, 4$. When the network enters a stable retrieval state for sufficiently high $\bar{m}_{\text{init}}$ below capacity, we shade the region and do not plot t_{opt} since no well-defined unique optimal time exists—any sufficiently late time yields the same high overlap.

In networks below capacity (with shaded regions), t_{opt} grows monotonically as $\bar{m}_{\text{init}}$ approaches the basin boundary. This growth becomes rapid and possibly divergent (though we cannot confirm this due to T being finite in our solutions) as $\bar{m}_{\text{init}}$ approaches the critical value for entering the basin of attraction.

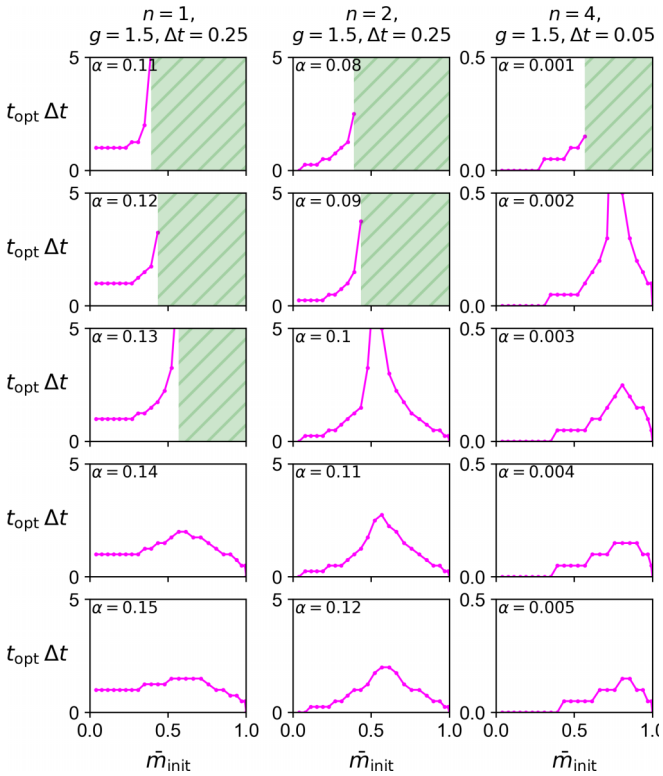

FIG. 5. Optimal readout time $t_{\text{opt}} = \arg\max_t \bar{m}(t)$ as a function of initial normalized overlap $\bar{m}_{\text{init}}$. Columns show $n = 1, 2, 4$ from left to right. Rows show increasing values of α. Magenta lines show DMFT values of t_{opt}. Green shaded regions indicate where the network enters a stable retrieval state, making the optimal readout time undefined since any late time works equally well.

In networks above capacity (without shaded regions), no basin of attraction exists and the optimal readout time curves become nonmonotonic. t_{opt} increases for small $\bar{m}_{\text{init}}$, reaches a maximum at intermediate values—roughly where the below-capacity critical $\bar{m}_{\text{init}}$ was located—then decreases back toward $t_{\text{opt}} = 0$ at $\bar{m}_{\text{init}} = 1$. This nonmonotonic behavior reflects the structure of the energy landscape: initial conditions with intermediate overlap can access the slow regions of the energy landscape near the stored pattern, making it worthwhile to wait as the system exploits these landscape features to increase overlap before eventual decay. In contrast, initial conditions with very high overlap are already near optimal alignment, while those with very low overlap cannot access the relevant landscape features. In these cases, there is little benefit to waiting, as the system either starts near its optimal retrieval performance or cannot meaningfully improve retrieval performance through transient dynamics.

V. DISCUSSION

A. Historical methods

While we have used the cavity method for its intuitive appeal and suitability to bipartite systems, most prior theoretical work on Hopfield dynamics has used path-integral approaches. We briefly review these methods, both to place our approach in context and to clarify the relationship between prior DMFT derivations and ours. One version of the path-integral approach, applicable to systems described by differential or difference equations (potentially with noise, e.g., Langevin dynamics), is known as the Martin–Siggia–Rose–De Dominicis–Janssen (MSRDJ) formalism [27–29]. This approach constructs a generating functional encoding correlation and response functions through its derivatives. The functional enforces the equations of motion using integral representations of delta functions. After averaging over the quenched disorder, order parameters are introduced that factor a resulting action across sites, and the order parameters are determined via saddle points at large N. This formalism has been used extensively to study the dynamics of disordered systems. For instance, Sompolinsky and Zippelius [30] used it to study Langevin equations for a soft-spin version of the Sherrington-Kirkpatrick model, and it has found rich applications in disordered recurrent neural networks [31–33]. The path-integral formalism also permits analysis of fluctuations around the $N \to \infty$ values of the order parameters, enabling computation of quantities such as the Lyapunov exponent [31] or the dimension of activity [21] in random recurrent neural networks.

For discrete spins evolving at finite temperature, the natural description uses master equations (e.g., describing Glauber dynamics), and path-integral approaches have been developed for these systems as well. Sommers [34] developed one such method for the Sherrington-Kirkpatrick model, although this derivation was later questioned [35]. Nevertheless, this formalism became the foundation for subsequent work on Hopfield dynamics. Rieger *et al.* [36] used the Sommers formulation to study Hopfield dynamics but did not solve the DMFT equations, instead recovering replica-symmetric results of [3,4] through long-time limits. A similar approach was taken by Horner *et al.* [37], who studied binary-spin Hopfield dynamics starting from a generating functional "derived from Langevin dynamics of soft spin variables or from Sommers's formulation," then solved the DMFT equations using approximation schemes. For a review of these methods, see Coolen and Sherrington [38] and, more recently, Coolen [39]. Another interesting work is that of Gardner *et al.* [40], who used essentially the MSRDJ formalism to study deterministic, discrete-time evolution of binary spins in the Hopfield model (as well as the Sherrington-Kirkpatrick model and, briefly, higher-order generalizations of the Hopfield model; see below), deriving DMFT equations and analytically examining the first few time steps of retrieval.

B. Prior work on dynamics of associative memory models

1. Hopfield model

As mentioned above, both Gardner and Horner had DMFT equations in hand, although their analyses had important limitations. Gardner *et al.* [40] derived full DMFT equations for the binary-spin Hopfield model with parallel dynamics but could only solve them analytically for the first few time steps, since the number of order parameters grows quadratically in time, rendering the analytics increasingly complicated.

Subsequent work sought to make progress by reducing the quadratic number of order parameters to a linear number at the cost of approximation. Horner *et al.* [37] started from

the full two-time DMFT equations and derived an approximation scheme that reduces them to evolution equations for single-time quantities, interpolating between exact short-time behavior and the replica-symmetric stationary solution. The resulting equations could be integrated numerically to obtain the full overlap trajectory $m(t)$, but the dynamics at intermediate times are approximate. Amari and Maginu [41] took a different approach, bypassing the DMFT entirely and instead appealing to heuristic Gaussianity assumptions about the noise experienced by each neuron. This yields evolution equations for a small number of macroscopic variables (the overlap and a variance), which also cannot capture the full two-time structure of the DMFT. Both works dealt with binary spins rather than the continuous variables studied here.

Both Horner and Gardner noted, but did not explore in detail, the nonmonotonic nature of retrieval that is central to our work. Gardner *et al.* [40] showed analytically that, for binary-spin Hopfield networks with parallel dynamics, the overlap with a stored pattern can increase during the first few time steps even above the critical capacity $\alpha_c \approx 0.14$, provided $\alpha < 2/\pi \approx 0.64$. Horner *et al.* [37] observed the same effect but described it as "not very efficient" and did not pursue it further. Finally, despite its heuristic nature, the work of Amari and Maginu [41] is notable as an early, explicit analysis of transient retrieval phenomenology.

2. Beyond Hopfield: Dense models

For dense associative memory models with higher-order interactions, Gardner *et al.* [40] provided the free energy as a function of order parameters to be determined via saddle points and computed the overlap after the first time step. However, the authors did not write out the full DMFT equations for these models, in contrast with the standard Hopfield case for which they derived the complete DMFT.

During preparation of the present paper, Mimura *et al.* [42] provided a derivation, using path-integral methods, of DMFT equations for binary-spin dense associative memory models with deterministic dynamics. Their calculation should agree with the one initiated by Gardner *et al.* [40]. In contrast with our work, they studied binary-spin rather than graded-activity networks, used discrete time ($\Delta t = 1$) rather than continuous time, and derived the DMFT using generating functionals rather than the cavity method. They did not attempt to solve the equations numerically; instead appealing to an approximation that eliminates the self-coupling in the single-site dynamics, which is distinct from the approximation strategies of Horner *et al.* [37] and Amari and Maginu [41] discussed above. They also did not analyze transient retrieval or energy dynamics, nor did they compare their results to simulations.

3. Beyond Hopfield: Vector models

Nicoletti *et al.* [43] studied analytically and in simulations a Hopfield model with d-dimensional vector spins [meaning that each spin is a vector confined to a $(d-1)$-sphere], focusing on how capacity and transient retrieval are affected by d. For $d = 1$, one recovers binary spins, and the first-step retrieval formula of Nicoletti *et al.* [43] recovers the corresponding result of Gardner *et al.* [40], namely, that transient retrieval occurs for $\alpha < 2/\pi$. For $d > 1$, Nicoletti *et al.* [43] showed that the critical capacity for equilibrium retrieval shrinks with d ($\alpha_c \propto 1/d$), while the critical capacity for first-step transient denoising grows with d ($\tilde{\alpha} \propto d$). Nicoletti *et al.* [43] note that they "do not know why the first-step retrieval phase grows with spin dimension d" and leave this as an open question, referencing the present paper's energy-landscape interpretation as a potential "geometric memory hypothesis." We propose that increasing spin dimension enlarges the slow regions of the energy landscape near each pattern, which would simultaneously account for both the reduced static capacity (as enlarged slow regions of neighboring patterns interfere at lower loads) and the enhanced transient retrieval (as the slow regions become more prominent and persist to higher loads). Further probing this hypothesis is an interesting direction for future work.

4. More analytically tractable associative memory models

A rather different well-studied case of Hopfield model dynamics is that of randomly diluted networks, where a random mask is applied following Hebbian construction of the weights to zero out many matrix elements [44–47]. In the limit where each neuron receives K inputs with $K \to \infty$, $N \to \infty$, and $K/N \to 0$, the preactivations become Gaussian, allowing an analytical solution similar to that for chaotic random recurrent networks [48]. An interesting finding of Pereira-Obilinovic *et al.* [47] is that, in this setting, overloading the network is associated with the emergence of chaotic attractors in the vicinity of stored patterns, a consequence of the random sparsity. In the fully connected models we study, overloading instead leads to the usual spin-glass phase governed by an energy function, and we characterize the dynamics on the resulting complex landscape.

One analytically tractable model of associative memory dynamics is that of Ref. [49], in which all N neuronal variables are jointly constrained to a sphere (as opposed to the vector model of Nicoletti *et al.* [43], where each individual neuron is a d-dimensional vector on its own sphere). Interactions are four-way in this model, since pairwise interactions do not allow a retrieval phase. In this case, one can use generating functional (or, presumably, cavity) techniques to derive differential equations governing the evolution of the two-time order parameters. This approach is similar to that used for p-spin-glass models, whose DMFT and aging behavior were studied in Ref. [50].

C. Dynamical perspective

Equipped with the DMFT solutions derived and numerically solved in this work, we have taken a dynamical rather than an equilibrium perspective on associative memory models. While equilibrium analyses reveal the existence and stability of memory states, the dynamical view provides complementary insights into transient evolution, where much of the interesting memory retrieval actually occurs. We demonstrated that patterns can be transiently retrieved even when stable attractors no longer exist due to slow regions that persist in the energy landscape near stored patterns. The idea of using dynamics to probe energy landscape structure has precedent; Sompolinsky and colleagues [30,51] famously used finite-temperature dynamics of soft-spin-glass models to

uncover the ultrametric energy-landscape structure underlying the replica solution.

We introduced transient-recovery curves (Sec. IV E) as a tool for characterizing retrieval performance without requiring stable states, revealing that transient recall behavior changes gracefully as capacity increases rather than exhibiting an abrupt breakdown. Methods for characterizing or reverse-engineering recurrent neural networks whose computations rely on transient dynamics remain nascent [16] compared with fixed-point-based methods [14], and this area warrants further attention.

The optimal time t_{opt} before reading out a memory can be quite long, particularly when the system is initialized near the edge of a basin of attraction below capacity or near where this edge used to be above capacity (Sec. IV G). In this regime, multiple steps of recurrent dynamics contribute substantially to retrieval beyond the static information stored in the weights. Indeed, in very sparsely connected networks, multiple time steps are required simply for information to propagate across the network [52]. A separate but related question is whether recurrent application of the same weights differs fundamentally from a deep feedforward architecture with untied weights; Bauer *et al.* [53] recently compared the two for learning general functions, although not retrieval specifically.

D. Applications to other recurrent networks

The analytical and numerical DMFT tools developed here should be useful for analyzing other large recurrent neural networks where transient dynamics play important roles. For example, minor modifications of our equations and numerical techniques allow for determining order parameters in randomly connected recurrent neural networks with varying levels of reciprocal correlation, $\rho = \langle J_{ij}J_{ji}\rangle / \langle J_{ij}^2\rangle$. This problem was studied in simulations by Ref. [54], and the mean-field theory was solved under the assumption of stationary dynamics (dependence of order parameters only on $t - t'$) by Ref. [55]. A similar approach was applied to the random Lotka-Volterra model in ecology by Ref. [17]. For $\rho = 1$, there exists a Lyapunov function and nonstationary behavior is guaranteed; such a system is analogous to zero-temperature dynamics in the Sherrington-Kirkpatrick model [30]. An open question remains whether nonstationary behavior persists (rather than eventually equilibrating to a time-translation invariant state) only at $\rho = 1$, for all ρ above a critical $\rho_c > 0$, or for all $\rho > 0$.

E. Exponential models

The dense associative memory models we study achieve polynomial capacity, $P = O(N^n)$. A separate line of work has studied associative memory models with exponential capacity, which arise when the energy function involves an exponential, rather than monomial, nonlinearity applied to the overlaps [56–58]. These models have attracted significant interest due to their connection with the attention mechanism in transformers [57]; for a recent review, see Ref. [59]. The equilibrium properties of exponential-capacity models have been analyzed by Lucibello and Mézard [58] using the observation that the noise term in the neuronal input takes the form of a random-energy-model free energy (see also Ref. [60]). This analysis has been extended to structured patterns living on low-dimensional manifolds [61] (see also Sec. V G). As Lucibello and Mézard [58] note, "the full analytic computation of attraction basins requires following a trajectory in time. This is a complicated task, which has not been done in the standard Hopfield model, and which is beyond the reach of our method." Our DMFT framework addresses precisely this gap for both the Hopfield and higher-order polynomial-capacity cases, where the capacity scaling $P = O(N^n)$ permits a controlled treatment of the noise from uncondensed patterns via expansions in $1/\sqrt{N}$. Extending the DMFT to exponential-capacity models is an interesting open problem; it would likely require treating a random-energy-model inner problem within the dynamical equations.

F. From transient to stable retrieval

An intriguing question is whether transient retrieval can be stabilized into persistent retrieval through additional mechanisms. Gardner *et al.* [40] suggested exploiting transient retrieval in below-capacity networks to ultimately land the system in a basin of attraction: "It might be possible after a few time steps of parallel iteration to define a way of annealing into a metastable state closer to the pattern."

For above-capacity networks, where no basin of attraction exists, a concrete realization of a biologically motivated stabilizing mechanism has recently appeared. In particular, Del Gaudio *et al.* [62] leveraged the theory of coupled neuronal-synaptic dynamics of Clark and Abbott [63], in which Hebbian synaptic plasticity operates on a timescale comparable to neuronal dynamics (see also Ref. [64]), to show that such ongoing Hebbian plasticity can deform the energy landscape during transient evolution, creating a stable minimum where previously only a slow, but unstable, region existed. The mechanism is analogous to a ball on a trampoline: the network's activity transiently visits the vicinity of a stored pattern, as our results predict for the static-weight case, and the fast plasticity simultaneously reshapes the local energy landscape to trap the system there.

More broadly, the interplay between multiple dynamical timescales in memory systems is an interesting direction [65], and the stabilization of transient retrieval through fast synaptic plasticity, as demonstrated by Del Gaudio *et al.* [62], is one concrete example of what such interactions can achieve.

G. Structured patterns, generalization, and creativity

Most work on associative memory has aimed to retrieve only previously stored patterns, treating spurious states as failure modes. However, generalization requires the opposite, namely, that the network develop *new* attractors corresponding to patterns never seen during storage. Whether and how such generalization emerges depends on the structure of the stored data, and a growing body of equilibrium analyses has begun to map out the possibilities.

Alemanno *et al.* [66] introduced a "supervised" Hebbian learning rule in which the synaptic matrix is constructed from class-mean patterns rather than individual examples,

establishing an equivalence of the resulting system with restricted Boltzmann machines. When training examples are noisy versions of underlying centroids, a generalization phase emerges in which individual noisy examples do not form attractors, but the centroids do. Agliari *et al.* [67] considered a related setting using random-matrix theory, studying the eigenvalue distribution of the weight matrix built from noisy examples, relating spectral properties to retrieval and generalization. Mézard [68] analyzed Hopfield networks storing patterns that reside on a low-dimensional subspace using a bipartite cavity method, where generalization corresponds to attractors spanning the subspace. More recently, Kalaj *et al.* [69] studied a related model with richer structure, the random-features Hopfield model, where patterns are generated through a nonlinear projection of latent variables. In the above-capacity regime, the network develops attractors corresponding to previously unseen combinations of learned features, namely, uniform combinations of subsets of features. These would classically be considered spurious states, but here correspond to meaningful points on the latent manifold. Related behavior has been observed in the context of diffusion models [70–72]. For dense associative memory models with exponential nonlinearities, Achilli *et al.* [61] have studied capacity under a conceptually similar data manifold hypothesis [73]; for patterns on a linear manifold, they showed analytically that the entire manifold becomes the sole attractor. In each of these settings, what would classically be considered a spurious state is instead interpreted as a meaningful generalization of the training data.

A separate line of work has modified the learning rule itself to improve capacity or promote generalization. Fachechi *et al.* [74], Agliari *et al.* [75] introduced a "dreaming" construction for Hopfield network weights, defining a family of interaction matrices parametrized by a "dreaming time" t_d that interpolates between the Hebbian rule ($t_d = 0$) and the pseudo-inverse rule ($t_d \to \infty$), increasing the storage capacity from $\alpha_c \approx 0.14$ to $\alpha_c \approx 1$. Agliari *et al.* [76] showed that gradient descent on a regularized loss for Hopfield networks yields this same family of interaction matrices, with t_d equal to the inverse of the regularization hyperparameter. This optimization perspective reveals that moderate regularization (intermediate t_d) enables generalization, with attractors forming at ground-truth centroids rather than individual noisy examples (similar to the supervised Hebbian setting of Alemanno *et al.* [66]), while excessive regularization (large t_d) causes overfitting. Serricchio *et al.* [77] introduced a distinct iterative algorithm called "daydreaming" that simultaneously reinforces stored patterns and erases spurious states at each learning step. On correlated data, it exploits correlations to increase storage capacity and basin sizes. Finally, D'Amico *et al.* [78] showed that maximizing a pseudolikelihood at zero temperature naturally produces an associative memory that, for structured datasets, transitions from memorization to generalization as training examples increase, developing attractors for unseen data. In a related setting, D'Amico and Negri [79] trained a recurrent self-attention network via pseudolikelihood and observed that both training and test examples appear as transient states of the dynamics.

All of the analyses described above are equilibrium or static in nature. Extending the dynamical analysis of the present paper to these structured and learned settings is a natural open direction. On structured data, the transient dynamics studied here may take on new significance: rather than transiently visiting the vicinity of a stored pattern before being driven away, the system might transiently explore "generalization regions" (centroids, feature mixtures, or latent manifolds) that reflect the underlying data structure. The methods developed here could reveal how networks navigate structured energy landscapes during retrieval, how dynamics interact with hierarchical or compositional structure in stored data, and whether transient retrieval properties are qualitatively different when the landscape has been shaped by learning rules that promote generalization.

More speculatively, one might view creative thought as a dynamical process on a complex energy landscape shaped by past experience, where the goal is not to converge to a minimum corresponding to a previously stored memory, but to explore the landscape's structure in novel ways.

H. Outlook

We have considered noiseless dynamics throughout this work. Understanding how noise affects transient retrieval is an interesting direction. Reference [80] showed that temporal correlations in noise can improve retrieval in the higher-order spherical generalization of the Hopfield model of Ref. [49].

Finally, neuroscience experiments could attempt to test whether biological neural networks exploit transient dynamics for memory retrieval, as our results suggest they could. Experimental signatures of this phenomenon could be detected through analysis of neural population activity during memory tasks. Concretely, one could compute time-varying similarity measures between population activity and stored memory patterns, analogous to our overlap function $m(t)$. The presence of transient retrieval would manifest as initial increases in pattern similarity followed by slower decay. Observing this in error trials would be a particularly compelling connection to behavior.

ACKNOWLEDGMENTS

The author thanks Jacob Zavatone-Veth, Blake Bordelon, L. F. Abbott, Ken Miller, Stefano Fusi, Ashok Litwin-Kumar, and Haim Sompolinsky for valuable discussions, comments, suggestions, and pointers. The author was supported during preparation of this work by the Gatsby Charitable Foundation and the Kavli Foundation.

DATA AVAILABILITY

The data that support the findings of this article are openly available [81].

APPENDIX A: NUMERICALLY STABLE $\mathcal{F}(\phi)$

For $\phi(x) = \tanh(x)$, $\mathcal{F}(\phi)$ can be evaluated in a numerically stable manner using

$$\mathcal{F}(\phi) = \frac{1}{2}\log(1-\phi^2) + x\phi \tag{A1}$$

$$= \log 2 + x - \mathrm{SoftPlus}(2x) + x\phi, \tag{A2}$$

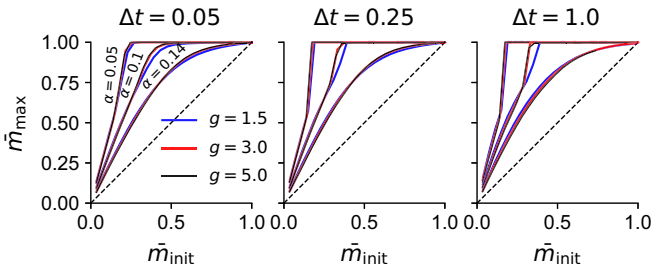

FIG. 6. Effects of gain g and time step Δt on transient retrieval dynamics for the Hopfield model ($n = 1$). Three panels from left to right show $\Delta t = 0.05, 0.25, 1.0$. Transient-recovery curves are shown for $\alpha = 0.05, 0.1, 0.14$, with gain values $g = 1.5, 3, 5$ distinguished by different colors for each α. Axes show maximum normalized overlap $\bar{m}_{\max}$ versus initial normalized overlap $\bar{m}_{\text{init}}$ as in Fig. 3. Curves for different gain values at the same α are closely grouped.

where $\tanh x = \phi$ and $\text{SoftPlus}(x) = \log(1 + e^x)$ has numerically stable implementations in common libraries.

APPENDIX B: FIXED-POINT MEAN-FIELD THEORY AND CRITICAL CAPACITIES

For completeness, we derive the fixed-point mean-field theory that yields the critical capacities reported in the main text. Equilibrium properties of the Hopfield model were computed using the replica method by Kühn *et al.* [18], and our results for $n = 1$ should agree with this analysis. Here, we obtain the fixed-point statistics for general n by taking the static limit of our DMFT cavity equations. This analysis confirms that the model reproduces the blackout catastrophe—the discontinuous vanishing of the nonzero overlap solution.

For a single condensed pattern, the equilibrium single-site equations result from removing time dependence from the DMFT equations, yielding the self-consistent system

$$x = \frac{g}{\sqrt{\alpha}}\xi m^n + \eta + F\phi(x), \tag{B1}$$

$$m = \langle \xi \phi(x) \rangle_{\text{single-site}}, \tag{B2}$$

$$C^\phi = \langle \phi^2(x) \rangle_{\text{single-site}}, \tag{B3}$$

$$S^\phi = \left\langle \frac{\phi'(x)}{1 - F\phi'(x)} \right\rangle_{\text{single-site}}, \tag{B4}$$

where we have dropped the subscript 0 for brevity and set $\sigma_\xi = 1$. The cavity field variance and self-coupling kernel are given by

$$C^\eta = \begin{cases} g^2 K^2 C^\phi & n = 1 \\ (2n-1)!!\, g^2 (C^\phi)^n & n > 1, \text{ even}, \end{cases} \tag{B5}$$

TABLE I. Critical capacities at $g = 1.5$.

Interaction order n	Critical capacity α_c (at $g = 1.5$)
1	0.13
2	0.080
4	0.0011

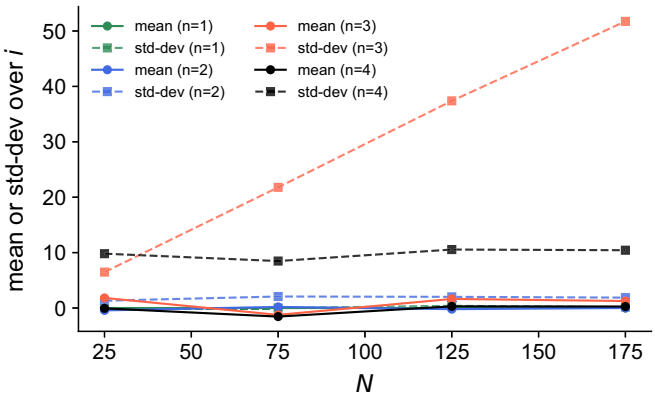

FIG. 7. Demonstration of divergent behavior for odd interaction orders. Mean and standard deviation of neuronal input $\text{input}_i = \frac{1}{\sqrt{\alpha}} \sum_\mu \xi_i^\mu (m^\mu)^n$ as functions of system size N for interaction orders $n = 1, 2, 3, 4$.

$$F = \begin{cases} g\sqrt{\alpha} K & n = 1 \\ n\,(2n-1)!!\, g^2 S^\phi (C^\phi)^{n-1} & n > 1, \text{ even}, \end{cases} \tag{B6}$$

where for the Hopfield case ($n = 1$), we have defined

$$K = \frac{1}{1 - \frac{g}{\sqrt{\alpha}} S^\phi}. \tag{B7}$$

The single-site average $\langle \cdot \rangle_{\text{single-site}}$ involves sampling $\xi \sim P(\xi)$ and $\eta \sim \mathcal{N}(0, C^\eta)$, then solving for x according to the nonlinear Eq. (B1).

With $\phi(x) = \tanh(x)$, an issue arises when $F \geqslant 1$, as the equation $x = F\phi(x) + A$ (where $A = \frac{g}{\sqrt{\alpha}}\xi m^n + \eta$) can have multiple solutions for a range of A values. This issue is resolved in the replica approach by requiring that the chosen solution minimizes a certain energy function [18]. There should be a way to recover this prescription in the cavity approach, making our equations agree with those of Kühn *et al.* [18] for $n = 1$ in our theory and temperature $\to 0$ in theirs, modulo self-interactions, but it is not immediately obvious how to do this.

To sidestep this multiple-solution issue, we restrict our analysis to $g = 1.5$, for which we find that F remains substantially below unity until the overlap solution m abruptly vanishes with increasing α. This allows us to compute the critical capacities where the high-overlap solution disappears (Table I).

APPENDIX C: EFFECTS OF g AND Δt

The transient retrieval phenomenology depends on the model parameters g and Δt. Figure 6 explores how varying these parameters affects transient retrieval through transient-recovery curves for the Hopfield model. Once g reaches ≈ 3 the curves largely converge. No clear trends are visible regarding Δt.

APPENDIX D: ODD n AND SCALING BEHAVIOR

To demonstrate the divergent behavior that occurs with odd n, we compute synthetic neuronal inputs, $\text{input}_i = \frac{1}{\sqrt{\alpha}} \sum_\mu \xi_i^\mu (m^\mu)^n$, where $m^\mu = \frac{1}{N} \sum_i \xi_i^\mu \phi_i$ and both ξ_i^μ and

ϕ_i are independently and identically distributed ± 1. We use $\alpha = 0.01$ and $P = \alpha N^n$. We compute this for each combination of $N \in \{25, 75, 125, 175\}$ and $n \in \{1, 2, 3, 4\}$. The mean and standard deviation of the neuronal input across the i index for a single draw of all variables are plotted in Fig. 7. With increasing N, the standard deviations remain flat (consistent with proper scaling) for all cases except $n = 3$, which grows linearly with N.

[1] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).

[2] J. J. Hopfield, Neurons with graded response have collective computational properties like those of two-state neurons, Proc. Natl. Acad. Sci. USA **81**, 3088 (1984).

[3] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Storing infinite numbers of patterns in a spin-glass model of neural networks, Phys. Rev. Lett. **55**, 1530 (1985).

[4] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Statistical mechanics of neural networks near saturation, Ann. Phys. (NY) **173**, 30 (1987).

[5] D. Tyulmankov, Computational models of learning and synaptic plasticity, in *Learning and Memory: A Comprehensive Reference*, 3rd ed., edited by J. T. Wixted (Elsevier, 2025).

[6] F. Zenke and A. Laborieux, Theories of synaptic memory consolidation and intelligent plasticity for continual learning, arXiv:2405.16922.

[7] L. F. Abbott and Y. Arian, Storage capacity of generalized networks, Phys. Rev. A **36**, 5091 (1987).

[8] E. Gardner, Multiconnected neural network models, J. Phys. A: Math. Gen. **20**, 3453 (1987).

[9] D. Horn and M. Usher, Capacities of multiconnected memory models, J. Phys. (Paris) **49**, 389 (1988).

[10] H. H. Chen, Y. C. Lee, G. Z. Sun, H. Y. Lee, T. Maxwell, and C. L. Giles, High order correlation model for associative memory, AIP Conf. Proc. **151**, 86 (1986).

[11] D. Krotov and J. J. Hopfield, Dense associative memory for pattern recognition, Adv. Neural Inf. Process. Syst. **29**, 1172 (2016).

[12] D. Krotov and J. J. Hopfield, Large associative memory problem in neurobiology and machine learning, in *The Ninth International Conference on Learning Representations* (ICLR, 2021).

[13] L. Kozachkov, J.-J. Slotine, and D. Krotov, Neuron–astrocyte associative memory, Proc. Natl. Acad. Sci. USA **122**, e2417788122 (2025).

[14] N. Maheswaranathan, A. H. Williams, M. D. Golub, S. Ganguli, and D. Sussillo, Universality and individuality in neural dynamics across large populations of recurrent networks, Adv. Neural Inf. Process. Syst. **32**, 15629 (2019).

[15] S. Vyas, M. D. Golub, D. Sussillo, and K. V. Shenoy, Computation through neural population dynamics, Annu. Rev. Neurosci. **43**, 249 (2020).

[16] E. Turner, K. V. Dabholkar, and O. Barak, Charting and navigating the space of solutions for recurrent neural networks, Adv. Neural Inf. Process. Syst. **34**, 25320 (2021).

[17] F. Roy, G. Biroli, G. Bunin, and C. Cammarota, Numerical implementation of dynamical mean field theory for disordered systems: Application to the Lotka–Volterra model of ecosystems, J. Phys. A: Math. Theor. **52**, 484001 (2019).

[18] R. Kühn, S. Bös, and J. L. van Hemmen, Statistical mechanics for networks of graded-response neurons, Phys. Rev. A **43**, 2084 (1991).

[19] D. Tyulmankov, C. Fang, A. Vadaparty, and G. R. Yang, Biological learning in key-value memory networks, Adv. Neural Inf. Process. Syst. **34**, 22247 (2021).

[20] A. van Meegen and H. Sompolinsky, Coding schemes in neural networks learning classification tasks, Nat. Commun. **16**, 3354 (2025).

[21] D. G. Clark, O. Marschall, A. Van Meegen, and A. Litwin-Kumar, Connectivity structure and dynamics of nonlinear recurrent neural networks, Phys. Rev. X **15**, 041019 (2025).

[22] D. Clark and H. Sompolinsky, Simplified derivations for high-dimensional convex learning problems, SciPost Phys. Lect. Notes 105 (2025).

[23] M. Mézard and G. Parisi, The cavity method at zero temperature, J. Stat. Phys. **111**, 1 (2003).

[24] M. Ramezanali, P. P. Mitra, and A. M. Sengupta, The cavity method for analysis of large-scale penalized regression, arXiv:1501.03194.

[25] J. W. Rocks and P. Mehta, Memorizing without overfitting: Bias, variance, and interpolation in overparameterized models, Phys. Rev. Res. **4**, 013201 (2022).

[26] M. Shamir and H. Sompolinsky, Thouless-Anderson-Palmer equations for neural networks, Phys. Rev. E **61**, 1839 (2000).

[27] P. C. Martin, E. D. Siggia, and H. A. Rose, Statistical dynamics of classical systems, Phys. Rev. A **8**, 423 (1973).

[28] C. De Dominicis, Dynamics as a substitute for replicas in systems with quenched random impurities, Phys. Rev. B **18**, 4913 (1978).

[29] J. A. Hertz, Y. Roudi, and P. Sollich, Path integral methods for the dynamics of stochastic and disordered systems, J. Phys. A: Math. Theor. **50**, 033001 (2017).

[30] H. Sompolinsky and A. Zippelius, Relaxational dynamics of the Edwards-Anderson model and the mean-field theory of spin-glasses, Phys. Rev. B **25**, 6860 (1982).

[31] A. Crisanti and H. Sompolinsky, Path integral approach to random neural networks, Phys. Rev. E **98**, 062120 (2018).

[32] M. Helias and D. Dahmen, *Statistical Field Theory for Neural Networks*, Lecture Notes in Physics, Vol. 970 (Springer, Cham, 2020).

[33] W. Zou and H. Huang, Introduction to dynamical mean-field theory of randomly connected neural networks with bidirectionally correlated couplings, SciPost Phys. Lect. Notes 79 (2024).

[34] H.-J. Sommers, Path-integral approach to Ising spin-glass dynamics, Phys. Rev. Lett. **58**, 1268 (1987).

[35] A. Lusakowski, Comment on "path-integral approach to Ising spin-glass dynamics", Phys. Rev. Lett. **66**, 2543 (1991).

[36] H. Rieger, M. Schreckenberg, and J. Zittartz, Glauber dynamics of the little-Hopfield model, Z. Phys. B: Condens. Matter **72**, 523 (1988).

[37] H. Horner, D. Bormann, M. Frick, H. Kinzelbach, and A. Schmidt, Transients and basins of attraction in neutral network models, Z. Phys. B: Condens. Matter **76**, 381 (1989).
[38] A. C. C. Coolen and D. Sherrington, Dynamics of attractor neural networks, in *Mathematical Approaches to Neural Networks*, edited by J. G. Taylor, North-Holland Mathematical Library, Vol. 51 (Elsevier, Amsterdam,1993), pp. 293–306.
[39] A. C. C. Coolen, Statistical mechanics of recurrent neural networks II—Dynamics, in *Handbook of Biological Physics* (Elsevier, Amsterdam, 2001), Vol. 4, pp. 619–684.
[40] E. Gardner, B. Derrida, and P. Mottishaw, Zero temperature parallel dynamics for infinite range spin glasses and neural networks, J. Phys. (Paris) **48**, 741 (1987).
[41] S.-I. Amari and K. Maginu, Statistical neurodynamics of associative memory, Neural Netw. **1**, 63 (1988).
[42] K. Mimura, J. Takeuchi, Y. Sumikawa, Y. Kabashima, and A. C. C. Coolen, Dynamical properties of dense associative memory, in *The Fourteenth International Conference on Learning Representations* (ICLR, 2026).
[43] F. Nicoletti, F. D'Amico, and M. Negri, Statistical mechanics of vector Hopfield network near and above saturation, J. Phys. A: Math. Theor. **58**, 505005 (2025).
[44] B. Derrida, E. Gardner, and A. Zippelius, An exactly solvable asymmetric neural network model, Europhys. Lett. **4**, 167 (1987).
[45] B. Derrida and J. Nadal, Learning and forgetting on asymmetric, diluted neural networks, J. Stat. Phys. **49**, 993 (1987).
[46] B. Tirozzi and M. Tsodyks, Chaos in highly diluted neural networks, Europhys. Lett. **14**, 727 (1991).
[47] U. Pereira-Obilinovic, J. Aljadeff, and N. Brunel, Forgetting leads to chaos in attractor networks, Phys. Rev. X **13**, 011009 (2023).
[48] H. Sompolinsky, A. Crisanti, and H.-J. Sommers, Chaos in random neural networks, Phys. Rev. Lett. **61**, 259 (1988).
[49] D. Bollé, T. M. Nieuwenhuizen, I. P. Castillo, and T. Verbeiren, A spherical Hopfield model, J. Phys. A: Math. Gen. **36**, 10269 (2003).
[50] A. Altieri, G. Biroli, and C. Cammarota, Dynamical mean-field theory and aging dynamics, J. Phys. A: Math. Theor. **53**, 375006 (2020).
[51] H. Sompolinsky, Time-dependent order parameters in spin-glasses, Phys. Rev. Lett. **47**, 935 (1981).
[52] D. Turcu and L. Abbott, Sparse RNNs can support high-capacity classification, PLoS Comput. Biol. **18**, e1010759 (2022).
[53] J. P. Bauer, K. Fischer, M. Helias, and A. Palmigiano, A unified theory of feature learning in RNNs and DNNs, arXiv:2602.15593.
[54] D. Martí, N. Brunel, and S. Ostojic, Correlations between synapses in pairs of neurons slow down dynamics in randomly connected neural networks, Phys. Rev. E **97**, 062314 (2018).
[55] D. G. Clark, L. F. Abbott, and A. Litwin-Kumar, Dimension of activity in random neural networks, Phys. Rev. Lett. **131**, 118401 (2023).
[56] M. Demircigil, J. Heusel, M. Löwe, S. Upgang, and F. Vermet, On a model of associative memory with huge storage capacity, J. Stat. Phys. **168**, 288 (2017).
[57] 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 *The Ninth InternationalConference on Learning Representations* (ICLR, 2021).
[58] C. Lucibello and M. Mézard, Exponential capacity of dense associative memories, Phys. Rev. Lett. **132**, 077301 (2024).
[59] D. Krotov, B. Hoover, P. Ram, and B. Pham, Modern methods in associative memory, arXiv:2507.06211.
[60] J. A. Zavatone-Veth and C. Pehlevan, Nadaraya–Watson kernel smoothing as a random energy model, J. Stat. Mech.: Theory Exp. (2025) 013404.
[61] B. Achilli, L. Ambrogioni, C. Lucibello, M. Mezard, and E. Ventura, The capacity of modern Hopfield networks under the data manifold hypothesis, in *New Frontiers in Associative Memories* (ICLR, 2025).
[62] M. Del Gaudio, F. Ghimenti, and S. Ganguli, Short-term plasticity recalls forgotten memories through a trampoline mechanism, arXiv:2511.22848.
[63] D. G. Clark and L. F. Abbott, Theory of coupled neuronal-synaptic dynamics, Phys. Rev. X **14**, 021001 (2024).
[64] J. Ba, G. E. Hinton, V. Mnih, J. Z. Leibo, and C. Ionescu, Using fast weights to attend to the recent past, Adv. Neural Inf. Process. Syst. **29**, 4331 (2016).
[65] A. J. Wakhloo, D. G. Clark, and L. F. Abbott, Associative synaptic plasticity creates dynamic persistent activity, bioRxiv (2025).
[66] F. Alemanno, M. Aquaro, I. Kanter, A. Barra, and E. Agliari, Supervised Hebbian learning, Europhys. Lett. **141**, 11001 (2023).
[67] E. Agliari, A. Fachechi, and D. Luongo, A spectral approach to Hebbian-like neural networks, Appl. Math. Comput. **474**, 128689 (2024).
[68] M. Mézard, Mean-field message-passing equations in the Hopfield model and its generalizations, Phys. Rev. E **95**, 022117 (2017).
[69] S. Kalaj, C. Lauditi, G. Perugini, C. Lucibello, E. M. Malatesta, and M. Negri, Random features Hopfield networks generalize retrieval to previously unseen examples, Phys. A (Amsterdam, Neth.) **678**, 130946 (2025).
[70] B. Pham, G. Raya, M. Negri, M. J. Zaki, L. Ambrogioni, and D. Krotov, Memorization to generalization: Emergence of diffusion models from associative memory networks, in *New Frontiers in Associative Memories* (ICLR, 2025).
[71] Z. Kadkhodaie, F. Guth, E. P. Simoncelli, and S. Mallat, Generalization in diffusion models arises from geometry-adaptive harmonic representations, in *The Twelfth International Conference on Learning Representations* (ICLR, 2024).
[72] M. Kamb and S. Ganguli, An analytic theory of creativity in convolutional diffusion models, in *Proceedings of the Forty-Second International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 267 (PMLR, 2025).
[73] S. Goldt, M. Mézard, F. Krzakala, and L. Zdeborová, Modeling the influence of data structure on learning in neural networks: The hidden manifold model, Phys. Rev. X **10**, 041044 (2020).
[74] A. Fachechi, E. Agliari, and A. Barra, Dreaming neural networks: Forgetting spurious memories and reinforcing pure ones, Neural Netw. **112**, 24 (2019).
[75] E. Agliari, F. Alemanno, A. Barra, and A. Fachechi, Dreaming neural networks: Rigorous results, J. Stat. Mech.: Theory Exp. (2019) 083503.

[76] E. Agliari, F. Alemanno, M. Aquaro, and A. Fachechi, Regularization, early-stopping and dreaming: A Hopfield-like setup to address generalization and overfitting, Neural Netw. **177**, 106389 (2024).

[77] L. Serricchio, D. Bocchi, C. Chilin, R. Marino, M. Negri, C. Cammarota, and F. Ricci-Tersenghi, Daydreaming Hopfield networks and their surprising effectiveness on correlated data, Neural Netw. **186**, 107216 (2025).

[78] F. D'Amico, D. Bocchi, L. M. Del Bono, S. Rossi, and M. Negri, Pseudo-likelihood produces associative memories able to generalize, even for asymmetric couplings, arXiv:2507.05147.

[79] F. D'Amico and M. Negri, Self-attention as an attractor network: Transient memories without backpropagation, in *2024 IEEE Workshop on Complexity in Engineering (COMPENG)* (IEEE, Piscataway, NJ, 2024), pp. 1–6.

[80] A. K. Behera, M. Rao, S. Sastry, and S. Vaikuntanathan, Enhanced associative memory, classification, and learning with active dynamics, Phys. Rev. X **13**, 041043 (2023).

[81] D. G. Clark, Code for "transient dynamics of associative memory models" (2025), GitHub repository, https://github.com/davidclark1/TransientDynamicsAssocMem.