Associative memory model with arbitrary Hebbian length

Zijian Jiang[*]
PMI Lab, School of Physics, Sun Yat-sen University, Guangzhou 510275, People's Republic of China

Jianwen Zhou[*]
*PMI Lab, School of Physics, Sun Yat-sen University, Guangzhou 510275, People's Republic of China
and CAS Key Laboratory for Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190,
People's Republic of China*

Tianqi Hou[*]
*Department of Physics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, People's Republic of China
and Theory Lab, Central Research Institute, 2012 Labs, Huawei Technologies Co., Ltd., Hong Kong Science Park, People's Republic of China*

K. Y. Michael Wong
Department of Physics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, People's Republic of China

Haiping Huang[†]
PMI Lab, School of Physics, Sun Yat-sen University, Guangzhou 510275, People's Republic of China

(Received 30 March 2021; accepted 29 November 2021; published 17 December 2021)

Conversion of temporal to spatial correlations in the cortex is one of the most intriguing functions in the brain. The learning at synapses triggering the correlation conversion can take place in a wide integration window, whose influence on the correlation conversion remains elusive. Here we propose a generalized associative memory model of pattern sequences, in which pattern separations within an arbitrary Hebbian length are learned. The model can be analytically solved, and predicts that a small Hebbian length can already significantly enhance the correlation conversion, i.e., the stimulus-induced attractor can be highly correlated with a significant number of patterns in the stored sequence, thereby facilitating state transitions in the neural representation space. Moreover, an anti-Hebbian component is able to reshape the energy landscape of memories, akin to the memory regulation function during sleep. Our work thus establishes the fundamental connection between associative memory, Hebbian length, and correlation conversion in the brain.

DOI: 10.1103/PhysRevE.104.064306

I. INTRODUCTION

Associative learning and memory is one fundamental brain function across many species including rodents and primates [1,2]. The standard Hopfield network, based on Hebbian learning rules, establishes a seminal model to explore rich properties of associative memory in both artificial and biological neural networks [3,4]. As a classic example, the monkey's temporal cortex was observed to be able to convert the temporal correlation of stimuli into the spatial correlation in neural activity [5,6], which can be modeled by considering Hebbian interactions among neighboring random independent patterns [7]. For an external stimulus being part of a temporally ordered sequence, the elicited neural activity has a correlation with neighboring patterns of the sequence which decays until vanishing at a finite separation of the patterns. This correlated attractor phase is in contrast to the Hopfield model where all attractors corresponding to the stored patterns are all uncorrelated fixed points in the network dynamics.

A recent study argued that the combination of Hebbian and anti-Hebbian learning can significantly increase the span of the temporal association [8]. However, the synaptic integration is still limited to neighboring patterns, like other previous works. Wide learning windows of various widths are ubiquitous in biological synaptic plasticity and contribute to sequence learning [9–11]. In particular, the shape of learning windows (e.g., the temporal interval of a spike-time-dependent plasticity at which the presynaptic and postsynaptic activity induce plasticity) is subject to neuromodulation and affects further sequence learning [11,12]. Whether this microscopic temporal correlation in synaptic learning affects the global behavior of correlated attractors remains therefore elusive. Hence, a full understanding of how the temporal correlation among stimuli evokes the spatially correlated neural activity particularly in a generalized associative memory model is important and yet still lacking.

Here we propose a theoretical model of associative memory with *arbitrary* Hebbian length, corresponding to wide learning windows. This model can be analytically solved, providing us exact mechanisms underlying the correlated attractor phase. In particular, we find that even with only Hebbian learning, the wide learning window can give rise to

[*] These authors contributed equally to this paper.
[†] huanghp7@mail.sysu.edu.cn

a large correlation span, which suggests a distinct synaptic mechanism from that argued in the recent work [8]. Most importantly, our model reveals that an anti-Hebbian learning for the nonconcurrent patterns could reshape the energy landscape, removing irrelevant attractors, which may connect to the hypothesis of unlearning effects in rapid-eye-movement sleep (e.g., getting rid of unimportant memories) [13–16].

II. MODEL

In this study, we construct an associative memory model by the Hebbian learning [17], which shapes the coupling strength between two neurons. We assume that all N neurons are fully connected without self-interactions, which constructs an associative memory of P random patterns ($\boldsymbol{\xi}$). These patterns form a cyclic sequence, corresponding to a repeated presentation of an ordered sequence of independent items in monkey experiments [5,6]. Therefore, the coupling matrix of the associative memory model can be specified as

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{P}\left[c\xi_i^\mu\xi_j^\mu + \gamma\sum_{r=1}^{d}\left(\xi_i^{\mu+r}\xi_j^\mu + \xi_i^\mu\xi_j^{\mu+r}\right)\right], \qquad (1)$$

where c specifies the standard Hebbian strength (concurrent Hebbian terms), γ specifies the coupling strength between r-separated patterns (nonconcurrent Hebbian terms), and d is thus the Hebbian length of our model. The case of $d = 1$ has been studied by previous works [7,8,18], while $d = 0$ recovers the standard Hopfield model [3,4,19]. Setting an arbitrary d corresponds to potential wide learning windows observed in neural circuits [9–11,20–22]. For simplicity, $P(\xi_i^\mu = \pm 1) = 1/2$ for each pair (i, μ).

The coupling is symmetric, and thus an equilibrium state $\mathbf{s}$ exists, captured by the Boltzmann distribution

$$P(\mathbf{s}) = \frac{1}{Z}e^{-\beta\mathcal{H}(\mathbf{s})}, \qquad (2)$$

where $\mathcal{H}(\mathbf{s}) = -\frac{1}{2}\sum_{i\neq j}J_{ij}s_is_j$ being the Hamiltonian, β is an inverse temperature, and Z is the pattern-dependent partition function. Note that we can rearrange the coupling matrix as $\mathbf{J} = \frac{1}{N}\boldsymbol{\xi}^{\mathrm{T}}\mathbf{X}\boldsymbol{\xi}$, where the circulant matrix $X_{\mu\nu}$ is introduced as follows [23]:

$$X_{\mu\nu} = c\delta_{\mu\nu} + \gamma\sum_{r=1}^{d}(\delta_{\mu,(\nu-r)\bmod P} + \delta_{\mu,(\nu+r)\bmod P}). \qquad (3)$$

Then the Hamiltonian can be expressed as $\mathcal{H}(\mathbf{s}) = -\frac{N}{2}\mathbf{m}^{\mathrm{T}}\mathbf{X}\mathbf{m}$, where $\mathbf{m}$ denotes the pattern-state overlap vector whose component $m_\mu = \frac{1}{N}\sum_i \xi_i^\mu s_i$.

Like in the standard Hopfield model [3,4], the binary state of each neuron is determined by its local field h_i, which can be written as $h_i = \sum_j J_{ij}s_j$. By inserting the coupling matrix, we get a new expression:

$$h_i = \sum_\mu \xi_i^\mu\left(cm_\mu + \gamma\sum_{r=1}^{d}(m_{\mu-r} + m_{\mu+r})\right). \qquad (4)$$

Due to the statistical independence of the patterns, the overlap has a mean-field expression in the limit $P/N \to 0$ (finite loading limit; see Appendix E),

$$m_\mu = \left\langle \xi^\mu \,\mathrm{sgn}\left[\sum_\nu \xi^\nu\left(cm_\nu + \gamma\sum_{r=1}^{d}(m_{\nu-r} + m_{\nu+r})\right)\right]\right\rangle, \qquad (5)$$

where $\langle\cdot\rangle$ denotes the disorder average over the pattern, and the zero-temperature limit ($\beta \to \infty$) is considered. In this limit, the dynamics is noiseless, and for $d = 1$ the overlap with the pattern used as a stimulus displays a largest value and was found to decay symmetrically until vanishing at a pattern-separation distance of five [7], which is independent of the number of patterns P (Fig. 1). This shows that, although the patterns are uncorrelated, the retrieved attractor starting from the stimulus has macroscopically significant overlaps with neighboring patterns within a finite distance. We call this kind of attractor a correlated attractor.

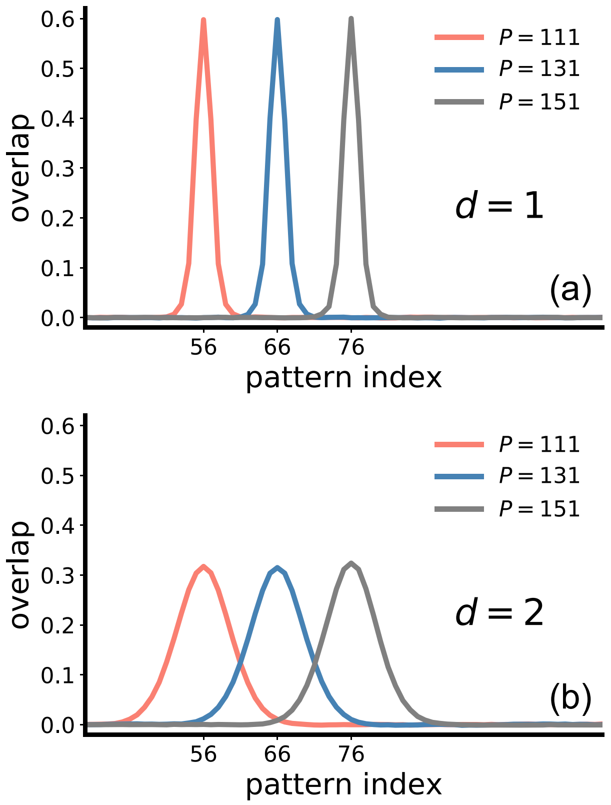

FIG. 1. The correlation span where the overlap peaking at the stimulus pattern decays to zero does not depend on the number of stored patterns. $c = 1$, and $\gamma = 1$. The pattern indexes 56, 66, and 76 are the stimulus pattern indexes corresponding to the cases of $P =$ 111, 131, and 151, respectively.

In the same spirit, the correlation of activities in two attractors can be computed as

$$\mathcal{C}(\mu, \mu') = \langle \mathrm{sgn}(h^\mu)\,\mathrm{sgn}(h^{\mu'})\rangle, \qquad (6)$$

where $h^\mu = \sum_\nu \xi^\nu(cm_\nu^\mu + \gamma\sum_{r=1}^{d}(m_{\nu-r}^\mu + m_{\nu+r}^\mu))$ (see Appendix E), and m_ν^μ defines the overlap of the attractor corresponding to the stimulus μ with the pattern number ν. The behavior of $\mathcal{C}(\mu, \mu')$ shows the emergence of correlated attractors from a network storing uncorrelated patterns. This attractor correlation decays with the separation of the patterns in the sequence from the stimulus pattern, where we can determine the critical distance (correlation length denoted as ℓ_c) beyond which the correlation value falls below 10^{-2}. This captures the basic coding strategy in the temporal cortex of the monkey, which is able to convert the temporal correlation among visual stimuli into a spatial correlation in the sustained neural activities evoked by the stimuli [5–7]. It is thus interesting to explore analytically how the Hebbian length (or other model parameters) affects properties of the correlated attractor.

III. A STATISTICAL MECHANICS ANALYSIS

Now we calculate the free energy of the model for the extensive-load case $\alpha = P/N \sim O(1)$. To derive a typical behavior of the model, we need to perform a disorder average of $\ln Z$, which can be tackled by the replica method: $-\beta f = \lim_{n\to 0, N\to\infty} \frac{\ln\langle Z^n\rangle}{nN}$ (e.g., see [24,25]). In essence, n copies of the original system are introduced. The analysis of the n replicas leads to the order parameters $m_\mu^a = \frac{1}{N}\sum_i^\mu \xi_i^\mu s_i^a$ and the state overlap $q_{ab} = \frac{1}{N}\sum_i s_i^a s_i^b$. For simplicity, we take the replica symmetric assumption [18], where the order parameters $(\{m_\mu^a, q_{ab}\})$ and their conjugate counterparts $(\{\hat m_\mu^a, \hat q_{ab}\})$ do not depend on the replica index (a or b). The final analytic form of the free energy reads as

$$\begin{aligned} -\beta f = {}& \frac{\beta^2 \hat q}{2}(q-1) - \frac{\beta}{2}\mathbf{m}^{\mathsf T}\mathbf{K}\mathbf{m} \\ & - \frac{\alpha}{2}\int_0^1 du \ln\left[1-\beta(1-q)\Lambda(u)\right] \\ & + \frac{\alpha\beta q}{2}\int_0^1 du \frac{\Lambda(u)}{1-\beta(1-q)\Lambda(u)} \\ & + \left\langle \int Dz \ln\left[2\cosh\left(\beta\sqrt{\hat q}z + \beta\boldsymbol{\xi}_F^{\mathsf T}\mathbf{K}\mathbf{m}\right)\right]\right\rangle, \end{aligned} \tag{7}$$

where the angular bracket denotes the disorder average over the condensed patterns $\boldsymbol{\xi}_F$, $\Lambda(u) = c + 2\gamma\sum_{r=1}^d \cos(2\pi r u)$ is the eigenvalue of the matrix $\mathbf{X}$ in the large P limit, $\mathbf{K}$ is an $S\times S$ matrix given by $\mathbf{F}_1 + C^{-1}\mathbb{1}$, where $\mathbb{1}$ is an identity matrix, and $(\mathbf{F}_1^{-1})_{ij} = w_{j-i}$ being a Toeplitz matrix [23], whose components (w_k) depend on both C and Λ. Hereafter, S denotes the number of condensed patterns (i.e., m_μ does not vanish as $N\to\infty$), and S can be larger than one due to the emergence of the correlated-attractor phase.

The calculation details are given in Appendixes A–C. In accord with the aforementioned noiseless dynamics, we are interested in the zero-temperature phase diagram. The finite-temperature analysis is straightforward (see Appendix B). In the zero-temperature limit, we denote $C \equiv \beta(1-q)$ as a finite order parameter, because $q\to 1$ in this limit.

The thermodynamic limit makes a saddle-point analysis of the free energy reasonable, which leads to the following saddle-point equations:

$$m_\mu = \left\langle \xi^\mu \,\mathrm{erf}\left(\frac{\sum_{\nu=1}^S \hat m_\nu \xi^\nu}{\sqrt{2\hat q}}\right)\right\rangle, \tag{8a}$$

$$\hat m_\mu = [\mathbf{K}\mathbf{m}]_\mu, \tag{8b}$$

$$C = \sqrt{\frac{2}{\pi\hat q}}\left\langle \exp\left[-\frac{\left(\sum_{\nu=1}^S \hat m_\nu \xi^\nu\right)^2}{2\hat q}\right]\right\rangle, \tag{8c}$$

$$\hat q = \alpha\int_0^1 du \frac{\Lambda(u)^2}{[1-C\Lambda(u)]^2} + \mathbf{m}^{\mathrm T}\frac{\partial\mathbf{K}}{\partial C}\mathbf{m}. \tag{8d}$$

The technical details for deriving these saddle-point equations are given in Appendix D. For the standard Hopfield model, $\mathbf{X} = \mathbb{1}$, $\Lambda(u) = 1$, and thus Eq. (8) reduces to the mean-field equation derived in the seminal work [19]. In our current setting, the Hebbian length affects both $\hat q$ and $\mathbf{K}$ in a highly nontrivial way. We thus expect the corresponding influence on the global behavior of the correlation conversion.

IV. RESULTS

We first study the mean-field dynamics [Eqs. (5) and (6)] of the overlap function at finite values of P, focusing on impacts of different model parameters. As shown in Fig. 2, increasing the Hebbian length lowers the peak value of the overlap with the stimulus pattern, and meanwhile, the overlap with neighboring patterns grows, thereby making the overlap profile broader. Surprisingly, by increasing the Hebbian length up to only $d = 2$, the correlation span is increased by quite a large margin (from $\ell_c = 5$ when $d = 1$ to $\ell_c = 15$ when $d = 2$). Compared to fine tuning the (negative) strength of the concurrent Hebbian terms [8], increasing the Hebbian length is simple and moreover biologically intuitive, as the Hebbian length corresponds to the size of the learning integration window, widely observed in neural circuits [9–11,20–22]. In particular, a large value of d allows for associations of patterns (stimuli) distant with each other in the sequence [Figs. 2(a) and 2(b)]. Furthermore, it requires only $d = 15$ for the correlation to expand to all patterns in the sequence, for $P = 151$ in Fig. 2(a). In other words, a small value of d can significantly amplify the correlation span [Fig. 2(c)]. The corresponding influence of d is tuned by the Hebbian strength γ, and a large value of γ has less impact on the tuning.

Therefore, our model with arbitrary Hebbian length provides a simple alternative way to control the correlation span of the stimulus-induced attractor, which is related to the conversion of the temporal correlations in the stored sequence into the spatial correlations of the neural activities. The correlated attractor phase is able to accelerate the transition between two highly correlated attractors (e.g., memories), since both attractors share a large number of common active neurons in their neural representations.

Next, we explore the effect of the nonconcurrent anti-Hebbian terms. These terms are characterized by negative values of γ, which competes with the concurrent Hebbian terms ($c > 0$). In addition, the anti-Hebbian terms correspond to the unlearning process introduced to verify the hypothesis of memory consolidation or erasure in sleep [13–15]. Here we find that the nonconcurrent anti-Hebbian terms remove some specified attractors, which appears in the original energy landscape of the model without anti-Hebbian effects. In contrast, the corresponding sign-reversed attractors are preferred, indicated by the negative overlaps in Fig. 2(d). This interesting observation could be explained by the energy landscape in terms of overlap functions. We recast the Hamiltonian as

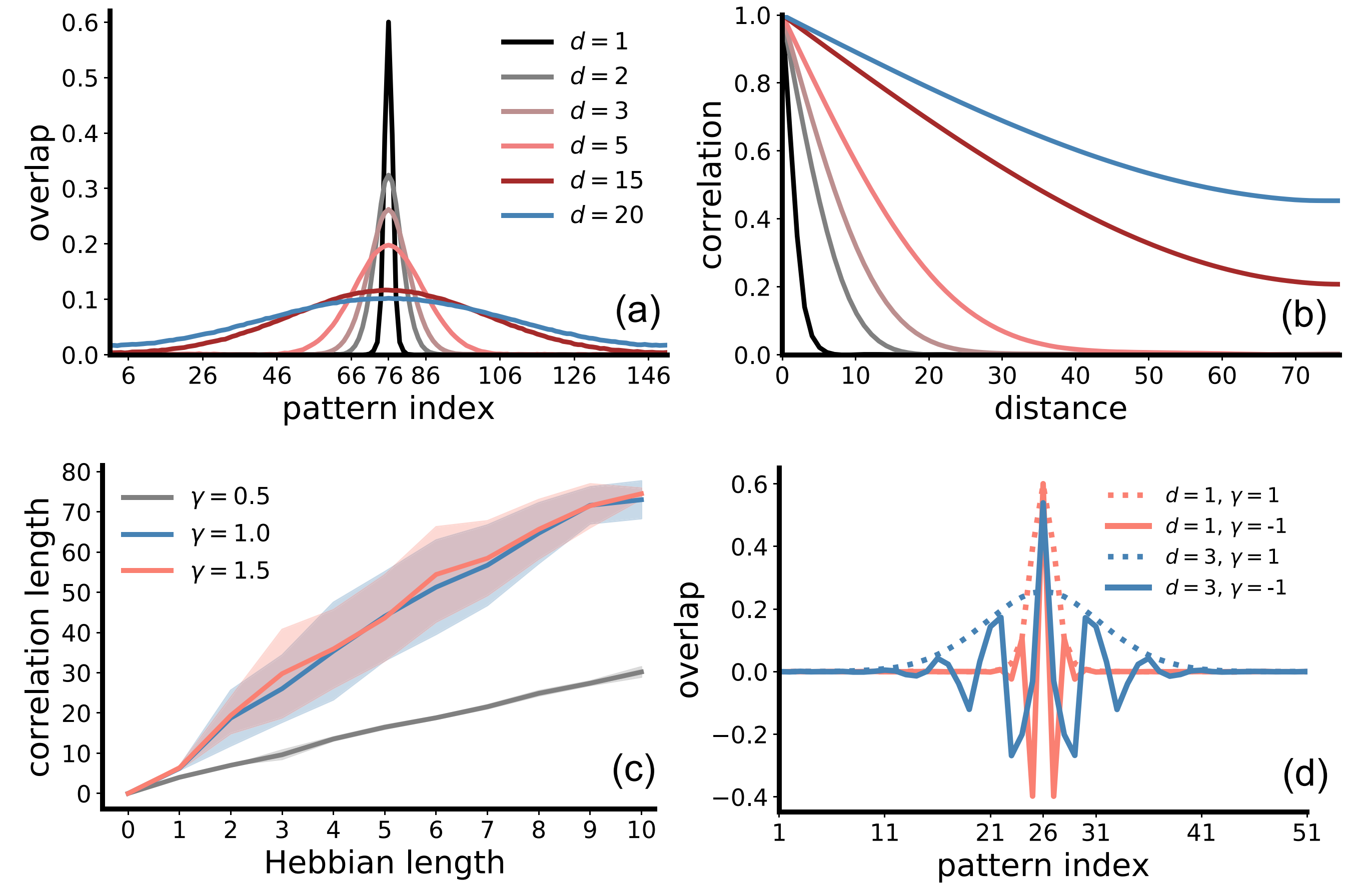

FIG. 2. Transforming temporal to spatial correlations with arbitrary Hebbian length. (a) Overlap profile with varying Hebbian length d. Other model parameters are $P = 151$, $c = 1.0$, and $\gamma = 1.0$. (b) Correlation between attractors vs their distance. The distance is defined as the separation from the corresponding stimulating patterns in the cyclic sequence. Other settings are the same as in (a). (c) Correlation length vs Hebbian length d. Other parameters are $P = 151$ and $c = 1.0$. The correlation length $\ell_c = \min\{\ell|\mathcal{C}(\ell = |\mu - \mu'|) < 10^{-2}\} - 1$. Fluctuations are the standard errors calculated from 30 trials. (d) Negative γ leads to the oscillatory behavior of the overlap profile. Other parameters are $P = 51$, and $c = 1.0$.

$\mathcal{H}(\mathbf{m}) = -\frac{Nc}{2}\sum_\mu (m_\mu)^2 - N\gamma \sum_\mu \sum_{r=1}^{d} m_\mu m_{\mu+r}$, where the first Hebbian term is always negative ($c > 0$), while the second term ($\gamma < 0$) requires that some specific overlap with a particular pattern index must take a negative value for a lower energy. A mathematical origin is that the gauge invariance in the coupling Eq. (1) is broken when $d > 1$, which is also reflected in the spectral density of the model [26]. In other words, the unlearning terms are able to reshape the energy landscape, by consolidating some memories while erasing other memories, akin to the function of both types of sleep: the rapid-eye-movement (REM) sleep is hypothesized to remove unnecessary memories while the slow-wave sleep contributes to the consolidation of important memories [15,27]. However, we remark that this connection must be further explored, e.g., in a recurrent circuit model with biological plasticity.

Interestingly, the overlap profile of $c = -1$ and $d = 1$ in statistical average is the same with that of $c = 1$ and $d = 2$ (Fig. 3), which implies that the effect of Hebbian length $d = 2$ on the correlated attractor is equivalent to the effect of anticoncurrent-Hebbian terms, provided that the energy $[\mathcal{H}(\mathbf{m}) = -\frac{N(c+2\gamma d)}{2}\sum_\mu (m_\mu)^2 + \frac{N\gamma}{2}\sum_\mu \sum_{r=1}^{d}(m_\mu - m_{\mu+r})^2]$ achieves a minimum in the correlated attractor phase. We thus argue that increasing the Hebbian length is an alternative way to expand the correlation span of each stimulus-induced attractor, for which an anti-Hebbian term [8] could be not necessary. It would be thus interesting to see if this prediction, despite being derived from the simplified

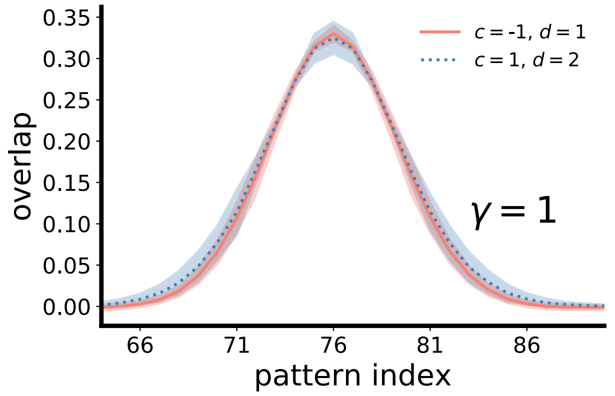

FIG. 3. Comparison of overlap profile of negative c and large d. The curves are the averages over 30 independent trials, and 10^7 Monte Carlo samples are used for running the mean-field dynamics. The shadowed region indicates the standard deviation.

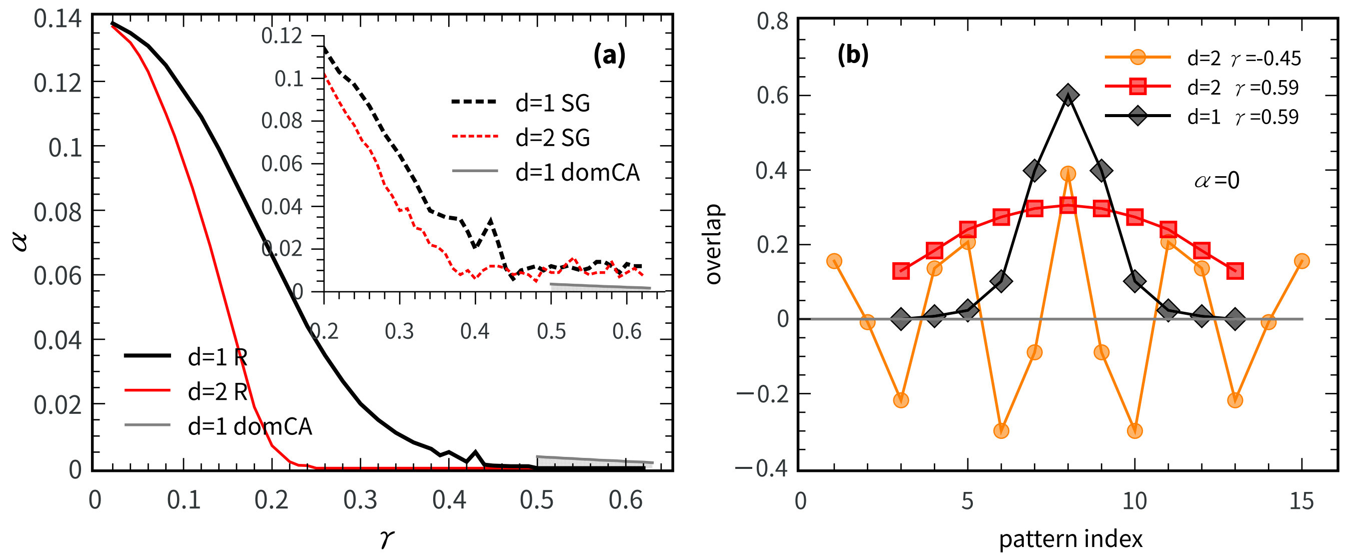

FIG. 4. Phase diagram of the associative memory model in the (α, γ) plane given $c = 1$. (a) The phase boundary shown by the lines delimits the retrieval (R) phase from the region where the correlated-attractor (CA) and spin-glass (SG) phases compete with each other (above the boundary). The boundary is the condition on which the retrieval phase loses its metastability from below. All shown transitions are of the discontinuous type. When $\alpha = 0$, the transition point is given by $\gamma_c = 0.5$ for $d = 1$, while $\gamma_c = 0.25$ for $d = 2$. The inset shows the boundary line above which the spin-glass phase is dominant. Note that for $d = 1$ there exists a very narrow regime (indicated by the shadow) within which the correlated-attractor phase is dominant (domCA). (b) Overlap profiles obtained from the statistical mechanics theory. All overlap profiles are defined as in Fig. 2, and obtained by solving the saddle-point equation of the model when $\alpha = 0$ and $d = 2$ (or $d = 1$). All theoretical results are obtained by assuming that $S = 11$, except that for negative values of γ, we use $S = 15$. Note that the results are not sensitive to the value of S (e.g., $S = 11$ or $S = 13$).

model, could be observed in biological circuits (e.g., in the hippocampus or temporal cortex).

Finally, we look at the phase diagram. We consider only $d = 1$ and $d = 2$. Other values of d could be analogously studied with our theory. As shown in Fig. 4(a), we identify three phases. One is the retrieval phase where only one overlap component is of the order one, i.e., $m_\mu = m\delta_{\mu\nu}$, where ν indicates the stimulating pattern. Given the value of α, increasing the value of γ would finally make the retrieval phase lose its metastability, after which the correlated-attractor phase becomes metastable. The line separating these two phases is thus the first-order transition. The correlated-attractor phase is characterized by the stimulus-induced attractors being highly correlated with a finite number of patterns in the stored sequence. In other words, the value of the corresponding overlap decays with the distance between the patterns in the sequence and the one used as the stimulus. The numerical solutions of the saddle-point equations obtained by the replica theory [Eq. (8)] reproduce the key features of the mean-field dynamics of the overlap [Fig. 2, and Fig. 4(b)], which corresponds to $\alpha = 0$ (finite loading) in our theory. Our theory thus predicts that the value of d can be used to expand the correlation span of the correlated attractor.

The Hebbian length could also reshape significantly the phase diagram. When $\alpha = 0$, the threshold for the dominant retrieval phase is $\gamma_c = 0.5$ for $d = 1$, but $\gamma_c = 0.25$ for $d = 2$. In the presence of a finite α, the retrieval phase loses its metastability at a smaller value of γ for $d = 2$ than for $d = 1$ [Fig. 4(a)]. After that, the spin-glass phase characterized by $m_\mu = 0$ $(\forall \mu)$ appears and competes with the correlated attractor phase, until the point where the spin-glass phase becomes dominant (global minimum of the free energy), as shown in the inset of Fig. 4(a). Remarkably, for $d = 1$, we identify a narrow regime for $\gamma > 0.5$ [the shadowed areas in Fig. 4(a)], where the correlated-attractor phase becomes dominant. This regime shrinks gradually as γ increases. For $d > 1$, this dominant phase is absent. The spin-glass phase always competes with the correlated-attractor phase until the spin-glass phase becomes dominant. One potential cause comes from the sum of cosine functions in the eigenvalue of the circulant matrix $\mathbf{X}$, which makes $\hat{q}$ fluctuate between small and large values.

If noisy neural dynamics is allowed (e.g., at a nonzero temperature), the spin-glass phase would be replaced by a paramagnetic phase at a continuous transition (see a detailed exploration in a companion paper [26]). This transition line is also strongly affected by the Hebbian length.

In particular, our theoretical analysis also reproduces the unlearning effects observed in the mean-field dynamics. Furthermore, a critical strength of $\gamma_c = -0.25$ for the oscillatory phase is predicted for $d = 2$, and $\gamma_c = -0.5$ for $d = 1$. When $\gamma < \gamma_c$, the unlearning effect of nonconcurrent anti-Hebbian terms takes place, preferring some particular patterns rather than their sign-reversed counterparts. In other words, the (spin reversal) symmetry in the Hamiltonian is broken, and the negative γ selects particular patterns, which suggests that the energy landscape is reshaped, and further the information storage is reoptimized, e.g., the storage capacity can be substantially improved by a local iterative unlearning procedure [28–31]. This intriguing phenomenon thus establishes the connection between the Hebbian length, anti-Hebbian effect, and memory function of unlearning.

V. CONCLUSIONS AND OUTLOOK

In this study, we propose the associative memory model of arbitrary Hebbian length, which considers both the wide learning integration window and temporal-to-spatial correlation conversion observed in the brain. Our theory predicts that a small value of Hebbian length (e.g., $d = 2$) can significantly expand the correlation span of the stimulus-induced attractors. Therefore, it seems unnecessary to fine tune the concurrent Hebbian strength c. Instead, by increasing d by only a small margin (e.g., from $d = 1$ to $d = 2$) can achieve the same goal of enhanced spatial correlations in neural attractors. Moreover, a negative value of γ can trigger an oscillatory behavior of the overlap profile, removing some irrelevant pattern attractors in the energy landscape, thereby playing the role of regulating the stored memories. Last, the Hebbian length could change strongly the phase diagram of the model. Increasing slightly the value of d would significantly suppress the retrieval phase, and moreover strongly affect the metastable regime of the correlation conversion. Taken together, our theory of the generalized associative memory model provides insights about the interplay between three important concepts—arbitrary Hebbian length, unlearning, and correlation conversion in neural networks.

Inspired by the recent work [8], if the pattern entry takes 0 or 1 (like being silent or emitting a spike), the coupling matrix can be decomposed into excitation and inhibition, $J_{ij} = J^{\mathrm{E}}_{ij} - J^{\mathrm{I}}_{ij}$, where

$$J^{\mathrm{E}}_{ij} = \frac{1}{N}\sum_{\mu=1}^{P}\left[c\xi_i^{\mu}\xi_j^{\mu} + \gamma\sum_{r=1}^{d}\left(\xi_i^{\mu+r}\xi_j^{\mu} + \xi_i^{\mu}\xi_j^{\mu+r}\right)\right], \tag{9a}$$

$$J^{\mathrm{I}}_{ij} = \frac{c}{NP}\sum_{\mu}\xi_i^{\mu}\sum_{\mu}\xi_j^{\mu} + \frac{\gamma}{PN}\sum_{r=1}^{d}\sum_{\mu}\xi_i^{\mu}\sum_{\mu}\xi_j^{\mu+r} + \frac{\gamma}{PN}\sum_{r=1}^{d}\sum_{\mu}\xi_j^{\mu}\sum_{\mu}\xi_i^{\mu+r}, \tag{9b}$$

from which we see clearly that both excitation and inhibition are modulated by the Hebbian length. It would then be interesting to explore how the memory attractor states are regulated by both excitation and inhibition, thereby establishing the relationship between long Hebbian interaction length and the timescale of episodic events during memory recall, which calls for future experimental studies of biological correlates of our theoretical predictions.

The encoding of pattern sequences in correlated attractors is reminiscent of encoding a continuous sequence of patterns in continuous attractor neural networks, which are useful for processing continuous information [32,33]. Our study may also help to understand how to link the synaptic plasticity with long temporal correlation to task-related activity (e.g., retrospective and prospective activity, related to a previously shown stimulus and a stimulus the monkey anticipate to appear, respectively) observed in pair association task experiments in monkeys [34,35].

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China, Grants No. 11805284 and 12122515 (H.H.), the start-up budget 74130-18831109 of the 100-Talent-Program of Sun Yat-sen University (H.H.), and the Research Grants Council of Hong Kong (Grants No. 16302419 and 16302619) (M.W.).

APPENDIX A: COMPUTATION OF THE DISORDER-AVERAGED FREE ENERGY

In our model, we specify the coupling matrix of neurons as

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{P}\left[c\xi_i^{\mu}\xi_j^{\mu} + \gamma\sum_{r=1}^{d}\left(\xi_i^{\mu}\xi_j^{\mu+r} + \xi_i^{\mu+r}\xi_j^{\mu}\right)\right], \tag{A1}$$

where ξ_i^{μ} follows independently a binary distribution, i.e., $p(\xi_i^{\mu} = \pm 1) = \frac{1}{2}\delta(\xi_i^{\mu} + 1) + \frac{1}{2}\delta(\xi_i^{\mu} - 1)$. We are interested in the limit of large P and N, thereby defining the ratio $\alpha = \frac{P}{N}$. α is also called the memory load of the associative memory model. Therefore, $\boldsymbol{\xi}$ is a $P \times N$ pattern matrix. The matrix $\mathbf{J}$ can be recast into the form

$$\mathbf{J} = \frac{1}{N}\boldsymbol{\xi}^{\mathrm{T}}\mathbf{X}\boldsymbol{\xi}, \tag{A2}$$

where $\mathbf{X}$ is a $P \times P$ circulant matrix, a special form of Toeplitz matrix with elements

$$\begin{aligned} X_{\mu\eta} &= c\delta_{\mu\eta} + \gamma\sum_{r=1}^{d}(\delta_{\mu,(\eta+r)\bmod P} + \delta_{\mu,(\eta-r)\bmod P}) \\ &= (c-\gamma)\delta_{\mu\eta} + \gamma\sum_{r=-d}^{d}\delta_{\mu,(\eta+r)\bmod P}. \end{aligned} \tag{A3}$$

The mth eigenvalue of $\mathbf{X}$ is given by [23]

$$\begin{aligned} \lambda_m &= \sum_{k=0}^{P-1}X_{1(k+1)}e^{-2\pi\mathrm{i}mk/P} \\ &= \sum_{k=0}^{P-1}X_{1(k+1)}\cos\left(2\pi\frac{mk}{P}\right) \\ &= \sum_{k=0}^{P-1}\left[c\delta_{0k} + \gamma\sum_{r=1}^{d}(\delta_{0,(k+r)\bmod P} + \delta_{0,(k-r)\bmod P})\right] \\ &\quad\times\cos\left(2\pi\frac{mk}{P}\right) \\ &= c + \gamma\sum_{r=1}^{d}\left[\cos\left(-2\pi\frac{mr}{P}\right) + \cos\left(2\pi\frac{mr}{P}\right)\right] \\ &= c + 2\gamma\sum_{r=1}^{d}\cos\left(2\pi\frac{mr}{P}\right), \end{aligned} \tag{A4}$$

for $m = 0, 1, \dots, P-1$.

The Hamiltonian of the model is defined by

$$\mathcal{H}(\mathbf{s}) = -\frac{1}{2}\sum_{i\neq j}J_{ij}s_is_j. \tag{A5}$$

The partition function is thus given by

$$Z = \mathrm{Tr}\exp\left(\frac{\beta}{2N}\mathbf{s}^{\mathsf{T}}\boldsymbol{\xi}^{\mathsf{T}}\mathbf{X}\boldsymbol{\xi}\mathbf{s}\right), \tag{A6}$$

where Tr indicates the summation over all discrete states $\mathbf{s} \equiv \{s_i = \pm 1\}_{i=1}^{N}$. To compute a disorder averaged free energy ($\langle -T\ln Z\rangle$) is in general computationally hard. However, the well-known replica trick developed in spin-glass theory [36] can be used to get around this difficulty, but assumptions on the replica matrix are required (detailed below). The replica method uses the mathematical identity

$$\langle \ln Z\rangle = \lim_{n\to 0}\frac{\ln\langle Z^n\rangle}{n}, \tag{A7}$$

where $\langle\cdot\rangle$ denotes the expectation over the distribution of $\boldsymbol{\xi}$. To proceed, we have to compute an integer power of the partition function:

$$Z^n = \mathrm{Tr}\exp\left[\frac{\beta}{2N}\sum_{a=1}^{n}(\mathbf{s}^a)^{\mathsf{T}}\boldsymbol{\xi}^{\mathsf{T}}\mathbf{X}\boldsymbol{\xi}\mathbf{s}^a\right], \tag{A8}$$

where Tr indicates the summation over all replicated states $\{\mathbf{s}^a\}_{a=1}^{n}$.

We consider the situation where there are S condensed (or foreground) patterns and $P-S$ noncondensed (or background) patterns, which is intuitive in our current setting. The choice of S can be justified *a posterior*, e.g., through solving the mean-field dynamics or saddle-point equations. Accordingly, we reorganize the matrix $\mathbf{X}$ in a form of block matrix,

$$\mathbf{X} = \begin{bmatrix}\mathbf{X}_{FF} & \mathbf{X}_{FB}\\ \mathbf{X}_{BF} & \mathbf{X}_{BB}\end{bmatrix}, \tag{A9}$$

where $\mathbf{X}_{FF}\in\mathbb{R}^{S\times S}$, $\mathbf{X}_{BF}^{\mathsf{T}} = \mathbf{X}_{FB}\in\mathbb{R}^{S\times(P-S)}$, and $\mathbf{X}_{BB}\in\mathbb{R}^{(P-S)\times(P-S)}$.

It then follows that

$$Z^n = \mathrm{Tr}\exp\left(\frac{\beta}{2N}\sum_{a,i,j,\mu\in B,\nu\in B}s_i^a\xi_i^\mu X_{\mu\nu}\xi_j^\nu s_j^a + \frac{\beta}{N}\sum_{a,i,j,\mu\in B,\nu\in F}s_i^a\xi_i^\mu X_{\mu\nu}\xi_j^\nu s_j^a + \frac{\beta}{2N}\sum_{a,i,j,\mu\in F,\nu\in F}s_i^a\xi_i^\mu X_{\mu\nu}\xi_j^\nu s_j^a\right). \tag{A10}$$

We then diagonalize the submatrix $\mathbf{X}_{BB}$ as $\mathbf{X}_{BB}^{\mu\nu} = \sum_\sigma \lambda_\sigma \eta_\mu^\sigma \eta_\nu^\sigma$, where λ_σ and η_μ^σ are denoted as the corresponding eigenvalues and eigenvectors, respectively. We thus obtain

$$\begin{aligned}Z^n &= \mathrm{Tr}\exp\left[\frac{\beta}{2N}\sum_{a,\sigma}\lambda_\sigma\left(\sum_{i,\mu\in B}s_i^a\xi_i^\mu\eta_\mu^\sigma\right)^2 + \frac{\beta}{N}\sum_{a,i,j,\mu\in B,\nu\in F}s_i^a\xi_i^\mu X_{\mu\nu}\xi_j^\nu s_j^a + \frac{\beta}{2N}\sum_{a,i,j,\mu\in F,\nu\in F}s_i^a\xi_i^\mu X_{\mu\nu}\xi_j^\nu s_j^a\right]\\ &= \mathrm{Tr}\prod_{a,\sigma}\int Dx_\sigma^a\exp\left[\sum_{i,\mu\in B}\frac{\xi_i^\mu}{\sqrt{N}}\left(\sum_{a,\sigma}s_i^a\eta_\mu^\sigma\sqrt{\beta\lambda_\sigma}x_\sigma^a + \frac{\beta}{\sqrt{N}}\sum_{a,j,\nu\in F}s_i^aX_{\mu\nu}\xi_j^\nu s_j^a\right) + \frac{\beta}{2N}\sum_{a,i,j,\mu\in F,\nu\in F}s_i^a\xi_i^\mu X_{\mu\nu}\xi_j^\nu s_j^a\right],\end{aligned} \tag{A11}$$

where we have used the Hubbard-Stratonovich transformation, i.e., $\exp(\frac{1}{2}b^2) = \int Dx\exp(\pm bx)$, where $Dx = \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})dx$.

We then define

$$\Phi_B = \exp\left[\sum_{i,\mu\in B}\frac{\xi_i^\mu}{\sqrt{N}}\left(\sum_{a,\sigma}s_i^a\eta_\mu^\sigma\sqrt{\beta\lambda_\sigma}x_\sigma^a + \frac{\beta}{\sqrt{N}}\sum_{a,j,\nu\in F}s_i^aX_{\mu\nu}\xi_j^\nu s_j^a\right)\right] \tag{A12}$$

and

$$\Phi_F = \exp\left[\frac{\beta}{2N}\sum_{a,i,j,\mu\in F,\nu\in F}s_i^a\xi_i^\mu X_{\mu\nu}\xi_j^\nu s_j^a\right]. \tag{A13}$$

Taking the disorder average over $\{\xi_i^\mu\}$, we write the result as

$$\langle Z^n\rangle = \left\langle \mathrm{Tr}\prod_{a,\sigma}\int Dx_\sigma^a\,\Phi_B\Phi_F\right\rangle. \tag{A14}$$

We first carry out the average over the distribution of background patterns, which yields

$$\langle\Phi_B\rangle = \exp\left\{\frac{1}{2N}\sum_{i,\mu\in B}\left[\sum_a s_i^a\left(\sum_\sigma\eta_\mu^\sigma\sqrt{\beta\lambda_\sigma}x_\sigma^a + \frac{\beta}{\sqrt{N}}\sum_{j,\nu\in F}X_{\mu\nu}\xi_j^\nu s_j^a\right)\right]^2\right\}. \tag{A15}$$

Introducing the state overlap as one order parameter, $q_{ab} = \frac{1}{N}\sum_i^N s_i^a s_i^b$ for $a \neq b$, and $m_\mu^a = \frac{1}{N}\sum_i \xi_i^\mu s_i^a$ as another order parameter, we have

$$
\begin{aligned}
\langle \Phi_B \rangle = \int \prod_{a\neq b} \frac{dq_{ab}\, d\hat{q}_{ab}}{2\pi/N} \prod_{a,\mu\in F} \frac{dm_\mu^a\, d\hat{m}_\mu^a}{2\pi/N} \exp\left[-\frac{1}{2}N\sum_{a\neq b}\hat{q}_{ab}q_{ab} + \frac{1}{2}\sum_{a\neq b}\hat{q}_{ab}\sum_i s_i^a s_i^b - N\sum_{a,\mu\in F} m_\mu^a \hat{m}_\mu^a + \sum_{a,\mu\in F}\hat{m}_\mu^a \sum_i \xi_i^\mu s_i^a\right] \\
\times \exp\left[\frac{1}{2}\sum_{\mu\in B}\sum_a \left(\sum_\sigma \eta_\mu^\sigma \sqrt{\beta\lambda_\sigma}x_\sigma^a + \beta\sqrt{N}\sum_{\nu\in F}X_{\mu\nu}m_\nu^a\right)^2\right] \\
\times \exp\left[\frac{1}{2}\sum_{\mu\in B}\sum_{a\neq b} q_{ab}\left(\sum_\sigma \eta_\mu^\sigma \sqrt{\beta\lambda_\sigma}x_\sigma^a + \beta\sqrt{N}\sum_{\nu\in F}X_{\mu\nu}m_\nu^a\right)\left(\sum_\sigma \eta_\mu^\sigma \sqrt{\beta\lambda_\sigma}x_\sigma^b + \beta\sqrt{N}\sum_{\nu\in F}X_{\mu\nu}m_\nu^b\right)\right],
\end{aligned}
\tag{A16}
$$

where an irrelevant factor $2^{-n(n-1)}$ is omitted.

Under the replica symmetric ansatz with $q_{ab} = q$ and $\hat{q}_{ab} = \hat{q}$ for $a \neq b$, $m_\mu^a = m_\mu$ and $\hat{m}_\mu^a = \hat{m}_\mu$, we arrive at

$$
\begin{aligned}
\langle \Phi_B \rangle = \int \frac{dq\, d\hat{q}}{(2\pi/N)^{n(n-1)}} \frac{dm\, d\hat{m}}{(2\pi/N)^{nS}} \exp\left[-\frac{1}{2}Nn(n-1)\hat{q}q + \frac{1}{2}\hat{q}\sum_{a\neq b}\sum_i s_i^a s_i^b - Nn\sum_{\mu\in F} m_\mu \hat{m}_\mu + \sum_{a,\mu\in F}\hat{m}_\mu \sum_i \xi_i^\mu s_i^a\right] \\
\times \exp\left[\frac{1}{2}\sum_{\mu\in B}\sum_a \left(\sum_\sigma \eta_\mu^\sigma \sqrt{\beta\lambda_\sigma}x_\sigma^a + \beta\sqrt{N}\sum_{\nu\in F}X_{\mu\nu}m_\nu\right)^2\right] \\
\times \exp\left[\frac{q}{2}\sum_{\mu\in B}\sum_{a\neq b} \left(\sum_\sigma \eta_\mu^\sigma \sqrt{\beta\lambda_\sigma}x_\sigma^a + \beta\sqrt{N}\sum_{\nu\in F}X_{\mu\nu}m_\nu\right)\left(\sum_\sigma \eta_\mu^\sigma \sqrt{\beta\lambda_\sigma}x_\sigma^b + \beta\sqrt{N}\sum_{\nu\in F}X_{\mu\nu}m_\nu\right)\right] \\
= \int \frac{dq\, d\hat{q}}{(2\pi/N)^{n(n-1)}} \frac{dm\, d\hat{m}}{(2\pi/N)^{nS}} \exp\left[-\frac{1}{2}Nn(n-1)\hat{q}q + \frac{1}{2}\hat{q}\sum_{a\neq b}\sum_i s_i^a s_i^b - Nn\sum_{\mu\in F} m_\mu \hat{m}_\mu + \sum_{a,\mu\in F}\hat{m}_\mu \sum_i \xi_i^\mu s_i^a\right] \\
\times \exp\left[\frac{1-q}{2}\sum_{\mu\in B}\sum_a \left(\sum_\sigma \eta_\mu^\sigma \sqrt{\beta\lambda_\sigma}x_\sigma^a + \beta\sqrt{N}\sum_{\nu\in F}X_{\mu\nu}m_\nu\right)^2\right] \exp\left[\frac{q}{2}\sum_{\mu\in B}\left(\sum_{a,\sigma} \eta_\mu^\sigma \sqrt{\beta\lambda_\sigma}x_\sigma^a + \beta n\sqrt{N}\sum_{\nu\in F}X_{\mu\nu}m_\nu\right)^2\right].
\end{aligned}
\tag{A17}
$$

We apply the Hubbard-Stratonovich transformation once again, and obtain

$$
\begin{aligned}
\langle \Phi_B \rangle = \int \frac{dq\, d\hat{q}}{(2\pi/N)^{n(n-1)}} \frac{dm\, d\hat{m}}{(2\pi/N)^{nS}} \prod_{\mu,a} Dy_\mu^a \prod_\mu Dz_\mu \exp\left[-\frac{1}{2}Nn(n-1)\hat{q}q + \frac{1}{2}\hat{q}\sum_{a\neq b}\sum_i s_i^a s_i^b\right. \\
\left. - Nn\sum_{\mu\in F} m_\mu \hat{m}_\mu + \sum_{a,\mu\in F}\hat{m}_\mu \sum_i \xi_i^\mu s_i^a\right] \exp\left[\sqrt{1-q}\sum_{\mu\in B}\sum_a \left(\sum_\sigma \eta_\mu^\sigma \sqrt{\beta\lambda_\sigma}x_\sigma^a + \beta\sqrt{N}\sum_{\nu\in F}X_{\mu\nu}m_\nu\right)y_\mu^a\right] \\
\times \exp\left[\sqrt{q}\sum_{\mu\in B}\left(\sum_{a,\sigma} \eta_\mu^\sigma \sqrt{\beta\lambda_\sigma}x_\sigma^a + \beta n\sqrt{N}\sum_{\nu\in F}X_{\mu\nu}m_\nu\right)z_\mu\right].
\end{aligned}
\tag{A18}
$$

By collecting all terms containing x_σ^a, we have

$$
\begin{aligned}
\langle \Phi_B \rangle = \int \frac{dq\, d\hat{q}}{(2\pi/N)^{n(n-1)}} \frac{dm\, d\hat{m}}{(2\pi/N)^{nS}} \prod_{\mu,a} Dy_\mu^a \prod_\mu Dz_\mu \exp\left[-\frac{1}{2}Nn(n-1)\hat{q}q + \frac{1}{2}\hat{q}\sum_{a\neq b}\sum_i s_i^a s_i^b\right. \\
\left. - Nn\sum_{\mu\in F} m_\mu \hat{m}_\mu + \sum_{a,\mu\in F}\hat{m}_\mu \sum_i \xi_i^\mu s_i^a\right] \exp\left[\sum_{a,\sigma} x_\sigma^a \sqrt{\beta\lambda_\sigma}\sum_{\mu\in B}\eta_\mu^\sigma\left(\sqrt{1-q}y_\mu^a + \sqrt{q}z_\mu\right)\right] \\
\times \exp\left[\beta\sqrt{N}\sum_{a,\mu\in B}\sum_{\nu\in F}X_{\mu\nu}m_\nu\left(\sqrt{1-q}y_\mu^a + \sqrt{q}z_\mu\right)\right].
\end{aligned}
\tag{A19}
$$

According to the definition of the overlap, Φ_F now can be written as

$$
\Phi_F = \exp\left[\frac{\beta n N}{2}\sum_{\mu\in F,\nu\in F} m_\mu X_{\mu\nu} m_\nu\right].
\tag{A20}
$$

Collecting all the results derived above, we have

$$\begin{aligned}\langle Z^n\rangle = {} & \mathrm{Tr}\int\prod_{a,\sigma}Dx_\sigma^a\frac{dq\,d\hat{q}}{(2\pi/N)^{n(n-1)}}\frac{dm\,d\hat{m}}{(2\pi/N)^{nS}}\prod_{\mu,a}Dy_\mu^a\prod_\mu Dz_\mu\exp\left[-\frac{1}{2}Nn(n-1)\hat{q}q+\frac{1}{2}\hat{q}\sum_{a\neq b}\sum_i s_i^a s_i^b-Nn\sum_{\mu\in F}m_\mu\hat{m}_\mu\right]\\ & \times\left\langle\exp\left[\sum_{a,\mu\in F}\hat{m}_\mu\sum_i\xi_i^\mu s_i^a\right]\right\rangle\exp\left[\sum_{a,\sigma}x_\sigma^a\sqrt{\beta\lambda_\sigma}\sum_{\mu\in B}\eta_\mu^\sigma\big(\sqrt{1-q}y_\mu^a+\sqrt{q}z_\mu\big)\right]\\ & \times\exp\left[\beta\sqrt{N}\sum_{a,\mu\in B}\sum_{\nu\in F}X_{\mu\nu}m_\nu\big(\sqrt{1-q}y_\mu^a+\sqrt{q}z_\mu\big)\right]\exp\left[\frac{\beta nN}{2}\sum_{\mu\in F,\nu\in F}m_\mu X_{\mu\nu}m_\nu\right].\end{aligned}\tag{A21}$$

We define the term summing over $\{s_i^a\}$ as

$$\begin{aligned}\langle\Phi_S\rangle &= \left\langle\mathrm{Tr}\exp\left[\frac{1}{2}\hat{q}\sum_{a\neq b}\sum_i s_i^a s_i^b+\sum_{a,\mu\in F}\hat{m}_\mu\sum_i\xi_i^\mu s_i^a\right]\right\rangle\\ &= \exp\left[-\frac{nN}{2}\hat{q}\right]\mathrm{Tr}\left\langle\prod_i\exp\left[\frac{1}{2}\hat{q}\left(\sum_a s_i^a\right)^2+\sum_{a,\mu\in F}\hat{m}_\mu\xi_i^\mu s_i^a\right]\right\rangle\\ &= \exp\left[-\frac{nN}{2}\hat{q}\right]\left\{\left\langle\mathrm{Tr}\exp\left[\frac{1}{2}\hat{q}\left(\sum_a s^a\right)^2+\sum_{a,\mu\in F}\hat{m}_\mu\xi^\mu s^a\right]\right\rangle\right\}^N.\end{aligned}\tag{A22}$$

Applying the Hubbard-Stratonovich transformation, we obtain

$$\begin{aligned}\langle\Phi_S\rangle &= \exp\left[-\frac{nN}{2}\hat{q}\right]\left\{\left\langle\int Dz\prod_a\mathrm{Tr}\exp\left[\sqrt{\hat{q}}s^a z+\sum_{\mu\in F}\hat{m}_\mu\xi^\mu s^a\right]\right\rangle\right\}^N\\ &= \exp\left[-\frac{nN}{2}\hat{q}\right]\left\{\left\langle\int Dz\prod_a 2\cosh\left[\sqrt{\hat{q}}z+\sum_{\mu\in F}\hat{m}_\mu\xi^\mu\right]\right\rangle\right\}^N\\ &= \exp\left[-\frac{nN}{2}\hat{q}\right]\exp\left\{N\ln\left[\left\langle\int Dz\,2^n\cosh^n\left(\sqrt{\hat{q}}z+\sum_{\mu\in F}\hat{m}_\mu\xi^\mu\right)\right\rangle\right]\right\}.\end{aligned}\tag{A23}$$

In the limit $n\to 0$,

$$\langle\Phi_S\rangle = \exp\left[-\frac{nN}{2}\hat{q}\right]\exp\left\{nN\left\langle\int Dz\,\ln\left[2\cosh\left(\sqrt{\hat{q}}z+\sum_{\mu\in F}\hat{m}_\mu\xi^\mu\right)\right]\right\rangle\right\}.\tag{A24}$$

Taken together, we have

$$\begin{aligned}\langle Z^n\rangle = {} & \int\prod_{a,\sigma}Dx_\sigma^a\frac{dq\,d\hat{q}}{(2\pi/N)^{n(n-1)}}\frac{dm\,d\hat{m}}{(2\pi/N)^{nS}}\prod_{\mu,a}Dy_\mu^a\prod_\mu Dz_\mu\\ & \times\exp\left[-\frac{1}{2}Nn(n-1)\hat{q}q-\frac{nN}{2}\hat{q}-Nn\sum_{\mu\in F}m_\mu\hat{m}_\mu+\frac{\beta nN}{2}\sum_{\mu\in F,\nu\in F}m_\mu X_{\mu\nu}m_\nu\right]\\ & \times\exp\left[\sum_{a,\sigma}x_\sigma^a\sqrt{\beta\lambda_\sigma}\sum_{\mu\in B}\eta_\mu^\sigma\big(\sqrt{1-q}y_\mu^a+\sqrt{q}z_\mu\big)\right]\\ & \times\exp\left[\beta\sqrt{N}\sum_{a,\mu\in B}\sum_{\nu\in F}X_{\mu\nu}m_\nu\big(\sqrt{1-q}y_\mu^a+\sqrt{q}z_\mu\big)\right]\\ & \times\exp\left\{nN\left\langle\int Dz\,\ln\left[2\cosh\left(\sqrt{\hat{q}}z+\sum_{\mu\in F}\hat{m}_\mu\xi^\mu\right)\right]\right\rangle\right\}.\end{aligned}\tag{A25}$$

To proceed, we first denote the vectors $\mathbf{y}^a=[y_\mu^a;\mu\in B]^\mathsf{T}$, $\mathbf{z}=[z_\mu;\mu\in B]^\mathsf{T}$, $\mathbf{m}=[m_\mu;\mu\in F]^\mathsf{T}$, $\hat{\mathbf{m}}=[\hat{m}_\mu;\mu\in F]^\mathsf{T}$, and $\boldsymbol{\xi}_F=[\xi^\mu;\mu\in F]^\mathsf{T}$. Integrating out $\{x_\sigma^a\}$, we get

$$\begin{aligned}
&\int\prod_{a,\sigma}Dx_\sigma^a\exp\left[\sum_{a,\sigma}x_\sigma^a\sqrt{\beta\lambda_\sigma}\sum_{\mu\in B}\eta_\mu^\sigma\big(\sqrt{1-q}y_\mu^a+\sqrt{q}z_\mu\big)\right]\\
&=\exp\left[\frac{1}{2}\beta\sum_{a,\sigma,\mu\in B,\nu\in B}\lambda_\sigma\eta_\mu^\sigma\eta_\nu^\sigma\big(\sqrt{1-q}y_\mu^a+\sqrt{q}z_\mu\big)\big(\sqrt{1-q}y_\nu^a+\sqrt{q}z_\nu\big)\right]\\
&=\exp\left[\frac{1}{2}\beta\sum_{a,\mu\in B,\nu\in B}X_{\mu\nu}(\sqrt{1-q}y_\mu^a+\sqrt{q}z_\mu)(\sqrt{1-q}y_\nu^a+\sqrt{q}z_\nu)\right]\\
&=\exp\left[\frac{1}{2}\beta(1-q)\sum_{a,\mu\in B,\nu\in B}y_\mu^aX_{\mu\nu}y_\nu^a+\beta\sqrt{(1-q)q}\sum_{a,\mu\in B,\nu\in B}z_\mu X_{\mu\nu}y_\nu^a+\frac{1}{2}n\beta q\sum_{\mu\in B,\nu\in B}z_\mu X_{\mu\nu}z_\nu\right]\\
&=\exp\left[\frac{1}{2}\beta(1-q)\sum_a(\mathbf{y}^a)^\mathsf{T}\mathbf{X}_{BB}\mathbf{y}^a+\beta\sqrt{(1-q)q}\sum_a\mathbf{z}^\mathsf{T}\mathbf{X}_{BB}\mathbf{y}^a+\frac{1}{2}n\beta q\mathbf{z}^\mathsf{T}\mathbf{X}_{BB}\mathbf{z}\right].
\end{aligned}\tag{A26}$$

Collecting all terms containing $\{y_\mu^a\}$, we get

$$\begin{aligned}
&\int\prod_{\mu,a}\frac{dy_\mu^a}{\sqrt{2\pi}}\prod_a\exp\left[-\frac{1}{2}\sum_{\mu\in B,\nu\in B}y_\mu^a[\delta_{\mu\nu}-\beta(1-q)X_{\mu\nu}]y_\nu^a\right]\\
&\quad\times\exp\left[\beta\sqrt{1-q}\sum_{\nu\in B}\left(\sum_{\mu\in F}\sqrt{N}X_{\nu\mu}m_\mu+\sqrt{q}\sum_{\mu\in B}z_\mu X_{\mu\nu}\right)y_\nu^a\right]\\
&=\int\prod_{\mu,a}\frac{dy_\mu^a}{\sqrt{2\pi}}\prod_a\exp\left[-\frac{1}{2}(\mathbf{y}^a)^\mathsf{T}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]\mathbf{y}^a\right]\\
&\quad\times\exp[\beta\sqrt{1-q}(\sqrt{N}\mathbf{m}^\mathsf{T}\mathbf{X}_{FB}+\sqrt{q}\mathbf{z}^\mathsf{T}\mathbf{X}_{BB})\mathbf{y}^a]\\
&=\frac{1}{\sqrt{[\det[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]]^n}}\exp\left[\frac{1}{2}n\beta^2(1-q)(\sqrt{N}\mathbf{m}^\mathsf{T}\mathbf{X}_{FB}+\sqrt{q}\mathbf{z}^\mathsf{T}\mathbf{X}_{BB})\right.\\
&\quad\left.\times[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}(\sqrt{N}\mathbf{X}_{BF}\mathbf{m}+\sqrt{q}\mathbf{X}_{BB}\mathbf{z})\right]\\
&=\frac{1}{\sqrt{[\det(\mathbb{1}-\beta(1-q)\mathbf{X}_{BB})]^n}}\exp\left[\frac{1}{2}nN\beta^2(1-q)\mathbf{m}^\mathsf{T}\mathbf{X}_{FB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BF}\mathbf{m}\right]\\
&\quad\times\exp\left[\frac{1}{2}n\beta^2(1-q)q\mathbf{z}^\mathsf{T}\mathbf{X}_{BB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BB}\mathbf{z}\right]\\
&\quad\times\exp[n\beta^2(1-q)\sqrt{Nq}\mathbf{m}^\mathsf{T}\mathbf{X}_{FB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BB}\mathbf{z}].
\end{aligned}\tag{A27}$$

We then collect all terms containing $\{z_\mu\}$, integrate out $\{z_\mu\}$ in the limit $n\to0$, and finally obtain

$$\begin{aligned}
&\int\prod_\mu\frac{dz_\mu}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\mathbf{z}^\mathsf{T}[\mathbb{1}-n\beta q\mathbf{X}_{BB}-n\beta^2(1-q)q\mathbf{X}_{BB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BB}]\mathbf{z}\right\}\\
&\quad\times\exp\{\beta n\sqrt{qN}[\mathbf{m}^\mathsf{T}\mathbf{X}_{FB}+\beta(1-q)\mathbf{m}^\mathsf{T}\mathbf{X}_{FB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BB}]\mathbf{z}\}\\
&=\exp\left\{-\frac{1}{2}\ln\det\big[\mathbb{1}-n\beta q\mathbf{X}_{BB}-n\beta^2(1-q)q\mathbf{X}_{BB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BB}\big]\right\},
\end{aligned}\tag{A28}$$

where to arrive at the last equality, we consider the limit of $n\to0$ [i.e., neglecting terms involving $O(n^2)$].

To sum up, we can write $\langle Z^n \rangle$ as

$$\begin{aligned}\langle Z^n\rangle = &\int \frac{dq\,d\hat{q}}{(2\pi/N)^{n(n-1)}}\prod_{\mu}\frac{dm_\mu\,d\hat{m}_\mu}{(2\pi/N)^{nS}} \times \exp\left\{nN\left\langle\int Dz\,\ln[2\cosh(\sqrt{\hat{q}}z+\hat{\mathbf{m}}^{\mathsf{T}}\boldsymbol{\xi}_F)]\right\rangle\right\}\\ &\times \exp\left[-\frac{1}{2}Nn(n-1)\hat{q}q-\frac{nN}{2}\hat{q}-Nn\mathbf{m}^{\mathsf{T}}\hat{\mathbf{m}}+\frac{\beta nN}{2}\mathbf{m}^{\mathsf{T}}\mathbf{X}_{FF}\mathbf{m}\right]\\ &\times \exp\left[-\frac{n}{2}\ln\det\,[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]\right]\times\exp\left[\frac{nN\beta^2(1-q)}{2}\mathbf{m}^{\mathsf{T}}\mathbf{X}_{FB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BF}\mathbf{m}\right]\\ &\times \exp\left[-\frac{1}{2}\ln\det\,[\mathbb{1}-n\beta q\mathbf{X}_{BB}-n\beta^2(1-q)q\mathbf{X}_{BB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BB}]\right].\end{aligned} \tag{A29}$$

By applying Laplace's method, we get the averaged free energy as

$$\begin{aligned}-\beta f \equiv \frac{1}{N}\langle\ln Z\rangle = &\left\langle\int Dz\,\ln[2\cosh(\sqrt{\hat{q}}z+\hat{\mathbf{m}}^{\mathsf{T}}\boldsymbol{\xi}_F)]\right\rangle+\frac{1}{2}\hat{q}q-\frac{1}{2}\hat{q}-\mathbf{m}^{\mathsf{T}}\hat{\mathbf{m}}+\frac{\beta}{2}\mathbf{m}^{\mathsf{T}}\mathbf{X}_{FF}\mathbf{m}\\ &-\frac{1}{2N}\ln\det\,[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]+\frac{\beta^2(1-q)}{2}\mathbf{m}^{\mathsf{T}}\mathbf{X}_{FB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BF}\mathbf{m}\\ &-\lim_{n\to 0}\frac{1}{2nN}\ln\det\,[\mathbb{1}-n\beta q\mathbf{X}_{BB}-n\beta^2(1-q)q\mathbf{X}_{BB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BB}].\end{aligned} \tag{A30}$$

This expression can be further simplified as

$$-\frac{1}{2N}\ln\det\,[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}] = -\frac{1}{2N}\sum_{\sigma}\ln\,[1-\beta(1-q)\lambda_\sigma] \tag{A31}$$

and

$$\begin{aligned}&-\lim_{n\to 0}\frac{1}{2nN}\ln\det\,[\mathbb{1}-n\beta q\mathbf{X}_{BB}-n\beta^2(1-q)q\mathbf{X}_{BB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BB}]\\ &\quad=-\lim_{n\to 0}\frac{1}{2nN}\sum_{\sigma}\ln\left[1-n\beta q\lambda_\sigma-\frac{n\beta^2(1-q)q\lambda_\sigma^2}{1-\beta(1-q)\lambda_\sigma}\right]\\ &\quad=\frac{1}{2N}\sum_{\sigma}\frac{\beta q\lambda_\sigma}{1-\beta(1-q)\lambda_\sigma}.\end{aligned} \tag{A32}$$

Finally, the averaged free energy is given by

$$\begin{aligned}\frac{1}{N}\langle\ln Z\rangle = &\left\langle\int Dz\,\ln[2\cosh(\sqrt{\hat{q}}z+\hat{\mathbf{m}}^{\mathsf{T}}\boldsymbol{\xi}_F)]\right\rangle+\frac{1}{2}\hat{q}q-\frac{1}{2}\hat{q}-\mathbf{m}^{\mathsf{T}}\hat{\mathbf{m}}+\frac{\beta}{2}\mathbf{m}^{\mathsf{T}}\mathbf{X}_{FF}\mathbf{m}-\frac{1}{2N}\sum_{\sigma}\ln\,[1-\beta(1-q)\lambda_\sigma]\\ &+\frac{\beta^2(1-q)}{2}\mathbf{m}^{\mathsf{T}}\mathbf{X}_{FB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BF}\mathbf{m}+\frac{1}{2N}\sum_{\sigma}\frac{\beta q\lambda_\sigma}{1-\beta(1-q)\lambda_\sigma}.\end{aligned} \tag{A33}$$

We rescale $\hat{q}$ by $\beta^2\hat{q}$, and $\hat{\mathbf{m}}$ by $\beta\hat{\mathbf{m}}$, and moreover define $\mathbf{K}=\mathbf{X}_{FF}+\beta(1-q)\mathbf{X}_{FB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BF}$. The stationary condition of the free energy with respect to $\mathbf{m}$ implies that $\hat{\mathbf{m}}=\mathbf{K}\mathbf{m}$. Therefore, the free energy can be reorganized as

$$\begin{aligned}-\beta f = &\frac{\beta^2\hat{q}}{2}(q-1)-\frac{\beta}{2}\mathbf{m}^{\mathsf{T}}\mathbf{K}\mathbf{m}-\frac{\alpha}{2}\int_0^1 du\,\ln\,[1-\beta(1-q)\Lambda(u)]+\frac{\alpha\beta q}{2}\int_0^1 du\frac{\Lambda(u)}{1-\beta(1-q)\Lambda(u)}\\ &+\left\langle\int Dz\,\ln\left[2\cosh\left(\beta\sqrt{\hat{q}}z+\beta\boldsymbol{\xi}_F^{\mathsf{T}}\mathbf{K}\mathbf{m}\right)\right]\right\rangle,\end{aligned} \tag{A34}$$

where $\Lambda(u)=c+2\gamma\sum_{r=1}^{d}\cos(2\pi ru)$. In the limit $P\to\infty$, it can be proved that X_{BB} is asymptotically equivalent to $\mathbf{X}$ [23]. Therefore, the summation over σ can be replaced by an integral using the eigenvalue of the circulant matrix $\mathbf{X}$.

APPENDIX B: DERIVATION OF SADDLE-POINT EQUATIONS

The order parameter should take values optimizing the free energy function, leading to the saddle-point equation (SDE). The saddle-point equation of q is given by

$$q-1+\frac{1}{\beta\sqrt{\hat{q}}}\left\langle\int Dz\, z\tanh(\beta\sqrt{\hat{q}}z+\beta\hat{\mathbf{m}}^{\mathsf{T}}\boldsymbol{\xi}_F)\right\rangle=0, \tag{B1}$$

$$q-1+\left\langle\int Dz\,[1-\tanh^2(\beta\sqrt{\hat{q}}z+\beta\hat{\mathbf{m}}^{\mathsf{T}}\boldsymbol{\xi}_F)]\right\rangle=0, \tag{B2}$$

$$q=\left\langle\int Dz\,\tanh^2(\beta\sqrt{\hat{q}}z+\beta\hat{\mathbf{m}}^{\mathsf{T}}\boldsymbol{\xi}_F)\right\rangle. \tag{B3}$$

The saddle-point equation of $\mathbf{m}$ is given by

$$\mathbf{m}=\left\langle\boldsymbol{\xi}_F\int Dz\,\tanh(\beta\sqrt{\hat{q}}z+\beta\hat{\mathbf{m}}^{\mathsf{T}}\boldsymbol{\xi}_F)\right\rangle. \tag{B4}$$

The saddle-point equation of $\hat{\mathbf{m}}$ is given by

$$\begin{aligned}\hat{\mathbf{m}}&=\mathbf{X}_{FF}\mathbf{m}+\beta(1-q)\mathbf{X}_{FB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BF}\mathbf{m}\\&:=\mathbf{K}\mathbf{m},\end{aligned} \tag{B5}$$

where $\mathbf{K}=\mathbf{X}_{FF}+\beta(1-q)\mathbf{X}_{FB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-1}\mathbf{X}_{BF}$, as derived at the end of the previous section. Finally, the saddle-point equation of $\hat{q}$ is given by

$$\begin{aligned}\hat{q}&=\frac{1}{N}\sum_{\sigma}\frac{q\lambda_\sigma^2}{[1-\beta(1-q)\lambda_\sigma]^2}\\&\quad+\mathbf{m}^{\mathsf{T}}\mathbf{X}_{FB}[\mathbb{1}-\beta(1-q)\mathbf{X}_{BB}]^{-2}\mathbf{X}_{BF}\mathbf{m}\\&=\alpha q\int_0^1\frac{\Lambda^2(u)\,du}{[1-\beta(1-q)\Lambda(u)]^2}-\beta^{-1}\mathbf{m}^{\mathsf{T}}\frac{\partial\mathbf{K}}{\partial q}\mathbf{m}.\end{aligned} \tag{B6}$$

Finally, the saddle-point equations are summarized as follows:

$$\hat{\mathbf{m}}=\mathbf{K}\mathbf{m}, \tag{B7a}$$

$$q=\left\langle\int Dz\,\tanh^2(\beta\sqrt{\hat{q}}z+\beta\hat{\mathbf{m}}^{\mathsf{T}}\boldsymbol{\xi}_F)\right\rangle, \tag{B7b}$$

$$\hat{q}=\alpha q\int_0^1 du\frac{\Lambda^2(u)}{[1-\beta(1-q)\Lambda(u)]^2}-\beta^{-1}\mathbf{m}^{\mathsf{T}}\frac{\partial\mathbf{K}}{\partial q}\mathbf{m}, \tag{B7c}$$

$$\mathbf{m}=\left\langle\boldsymbol{\xi}_F\int Dz\,\tanh(\beta\sqrt{\hat{q}}z+\beta\hat{\mathbf{m}}^{\mathsf{T}}\boldsymbol{\xi}_F)\right\rangle. \tag{B7d}$$

We analyze the critical temperature between the paramagnetic phase and spin-glass phase. In the spin-glass phase, $q\neq 0$ but $\mathbf{m}=0$. Expanding $q=\langle\int Dz\tanh^2(\beta\sqrt{\hat{q}}z+\beta\hat{\mathbf{m}}^{\mathsf{T}}\boldsymbol{\xi}_F)\rangle$, and $\hat{q}=\alpha q\int_0^1 du\frac{\Lambda^2(u)}{[1-\beta(1-q)\Lambda(u)]^2}+\mathbf{m}^{\mathsf{T}}\frac{\partial\mathbf{K}}{\partial C}\mathbf{m}$ $[C\equiv\beta(1-q)]$ in powers of q and $\hat{q}$, we have

$$q\simeq\beta^2\hat{q}\simeq\beta^2\alpha q\int_0^1 du\frac{\Lambda^2(u)}{[1-\beta\Lambda(u)]^2}+O(q^2). \tag{B8}$$

T_g can be obtained by solving

$$1=\alpha\int_0^1 du\frac{\Lambda^2(u)}{[T_{\mathrm{g}}-\Lambda(u)]^2}. \tag{B9}$$

For the standard Hopfield model, Eq. (B9) can be analytically solved with the result $T_{\mathrm{g}}=1+\sqrt{\alpha}$.

APPENDIX C: A COMPUTATION TRANSFORMATION TO SOLVE SDE

To solve the SDE numerically is challenging, due to the computation of $\mathbf{K}$, which involves the block structure of $\mathbf{X}$. To get rid of dependence on N and P (we are interested in only the large N and P limit), we propose the following numerical technique. We first define $C=\beta(1-q)$.

Note that if $C=0$, $\mathbf{K}=\mathbf{X}_{FF}$, $\frac{\partial\mathbf{K}}{\partial C}=\mathbf{X}_{FB}\mathbf{X}_{BF}$. Let

$$\mathbf{X}\mathbf{X}^{\top}=\begin{bmatrix}\mathbf{H}&\cdots\\\cdots&\cdots\end{bmatrix}, \tag{C1}$$

where $\mathbf{H}$ is an $S\times S$ symmetric matrix. Then we have

$$\mathbf{H}=\mathbf{X}_{FF}\mathbf{X}_{FF}^{\top}+\mathbf{X}_{FB}\mathbf{X}_{BF}=\mathbf{X}_{FF}\mathbf{X}_{FF}^{\top}+\left.\frac{\partial\mathbf{K}}{\partial C}\right|_{C=0}. \tag{C2}$$

The matrix $\mathbf{H}$ can be computed as

$$\mathbf{H}=\begin{bmatrix}h_0&h_1&\cdots&h_{S-1}\\h_1&h_0&\cdots&h_{S-2}\\\vdots&\vdots&&\vdots\\h_{S-1}&h_{S-2}&\cdots&h_0\end{bmatrix}, \tag{C3}$$

where

$$\begin{aligned}h_l&=\frac{1}{P}\sum_{m=0}^{P-1}\left[c+2\gamma\sum_{r=1}^{d}\cos\left(\frac{2\pi rm}{P}\right)\right]^2\exp\left(\frac{2\pi\mathrm{i}ml}{P}\right)\\&=\int_0^1 dx\left[c+2\gamma\sum_{r=1}^{d}\cos(2\pi rx)\right]^2\cos(2\pi lx).\end{aligned} \tag{C4}$$

Finally, we arrive at

$$\left.\frac{\partial\mathbf{K}}{\partial C}\right|_{C=0}=\mathbf{H}-\mathbf{X}_{FF}\mathbf{X}_{FF}^{\top}=\mathbf{H}-(\mathbf{K}|_{C=0})^2, \tag{C5}$$

where we used the fact that when $C=0$, $\mathbf{K}=\mathbf{X}_{FF}$.

If $C\neq 0$, we have $\mathbf{K}=\mathbf{X}_{FF}-\mathbf{X}_{FB}\frac{1}{\mathbf{X}_{BB}-C^{-1}\mathbb{I}}\mathbf{X}_{BF}$. To calculate $\mathbf{K}$ numerically in the large P limit, we notice that

$$(\mathbf{X}-C^{-1}\mathbb{I})^{-1}=\begin{bmatrix}\mathbf{F}_1^{-1}&\cdots\\\cdots&\cdots\end{bmatrix}, \tag{C6}$$

where $\mathbf{F}_1^{-1}\in\mathbb{R}^{S\times S}$ and is a submatrix of $(\mathbf{X}-C^{-1}\mathbb{I})^{-1}$. Since $\mathbf{X}-C^{-1}\mathbb{1}$ is a circulant matrix, its inverse matrix can be calculated by $(\mathbf{X}-C^{-1}\mathbb{1})^{-1}=\mathrm{Circ}(w_0,w_1,\ldots,w_{P-1})$, where

$$w_k=\int_0^1 dx\frac{\cos(2\pi kx)}{c-C^{-1}+2\gamma\sum_{r=1}^{d}\cos(2\pi rx)}, \tag{C7}$$

for $k=0,1,\ldots,P-1$ in the limit $P\to\infty$. Thus $\mathbf{F}_1^{-1}$ can be written as

$$\mathbf{F}_1^{-1}=\begin{bmatrix}w_0&w_1&\cdots&w_{S-1}\\w_1&w_0&\cdots&w_{S-2}\\\vdots&\vdots&&\vdots\\w_{S-1}&w_{S-2}&\cdots&w_0\end{bmatrix}. \tag{C8}$$

By using the matrix formula for the inverse of a block matrix, we can prove that $\mathbf{K}$ can be expressed as

$$\mathbf{K} = \mathbf{F}_1 + C^{-1}\mathbb{1}. \tag{C9}$$

Thus, to calculate $\mathbf{K}$ numerically, we first calculate w_k for $k = 0, 1, \dots, S-1$ to get $\mathbf{F}_1^{-1}$, and then calculate its inverse matrix $\mathbf{F}_1$, and finally add the matrix $C^{-1}\mathbb{1}$ to $\mathbf{F}_1$.

The term $\frac{\partial \mathbf{K}}{\partial C} = -\frac{1}{\beta}\frac{\partial \mathbf{K}}{\partial q}$ can be calculated as

$$\frac{\partial \mathbf{K}}{\partial C} = \frac{\partial \mathbf{F}_1}{\partial C} - \frac{1}{C^2}\mathbb{1} = -\mathbf{F}_1\frac{\partial \mathbf{F}_1^{-1}}{\partial C}\mathbf{F}_1 - \frac{1}{C^2}\mathbb{1}, \tag{C10}$$

where the entry of $\frac{\partial \mathbf{F}_1^{-1}}{\partial C}$ is computed as

$$\frac{\partial w_k}{\partial C} = -\int_0^1 dx \frac{C^{-2}\cos(2\pi kx)}{\left[c - C^{-1} + 2\gamma\sum_{r=1}^{d}\cos(2\pi rx)\right]^2}, \tag{C11}$$

for $k = 0, 1, \dots, S-1$.

APPENDIX D: ZERO-TEMPERATURE LIMIT

In the limit $T \to 0$ ($\beta \to \infty$), it is easy to derive that

$$\begin{aligned}\int Dz \tanh[\beta(\sqrt{\hat{q}}z + x)] &= \sqrt{\frac{2}{\pi}}\int_0^{\frac{1}{\sqrt{\hat{q}}}x} dz \exp\left(-\frac{1}{2}z^2\right) + O(T)\\ &\equiv \operatorname{erf}\left(\frac{1}{\sqrt{2\hat{q}}}x\right) + O(T),\end{aligned} \tag{D1}$$

and

$$\begin{aligned}\int \frac{dz}{\sqrt{2\pi}}e^{-z^2/2}[1-\tanh^2\beta(az+b)] &\simeq \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\Big|_{\tanh^2\beta(az+b)=0} \times \int dz[1-\tanh^2\beta(az+b)]\\ &= \frac{1}{\sqrt{2\pi}}e^{-b^2/2a^2}\frac{1}{a\beta}\int dz\frac{\partial}{\partial z}\tanh\beta(az+b)\\ &= \sqrt{\frac{2}{\pi}}\frac{1}{a\beta}e^{-b^2/2a^2}.\end{aligned} \tag{D2}$$

We thus obtain

$$\mathbf{m} = \left\langle \boldsymbol{\xi}_F \operatorname{erf}\left(\frac{1}{\sqrt{2\hat{q}}}\boldsymbol{\xi}_F^{\mathsf{T}}\mathbf{K}\mathbf{m}\right)\right\rangle. \tag{D3}$$

In the limit $T \to 0$, we also have

$$\begin{aligned}\beta(1-q) &= \beta\int Dz\left\langle 1 - \tanh^2\left(\beta\sqrt{\hat{q}}z + \beta\boldsymbol{\xi}_F^{\mathsf{T}}\mathbf{K}\mathbf{m}\right)\right\rangle\\ &= \sqrt{\frac{2}{\pi\hat{q}}}\left\langle \exp\left[-\frac{\left(\boldsymbol{\xi}_F^{\mathsf{T}}\mathbf{K}\mathbf{m}\right)^2}{2\hat{q}}\right]\right\rangle\\ &\equiv C.\end{aligned} \tag{D4}$$

The conjugated order parameter $\hat{q}$ is given by

$$\hat{q} = \alpha\int_0^1 du\frac{\Lambda^2(u)}{[1-C\Lambda(u)]^2} + \mathbf{m}^{\mathsf{T}}\frac{\partial \mathbf{K}}{\partial C}\mathbf{m}, \tag{D5}$$

where we have used the fact that in the zero-temperature limit $q \to 1$.

The free energy at the zero-temperature limit is given by

$$\begin{aligned}-f = {}&\frac{\alpha}{2}\int_0^1 du\frac{\Lambda(u)}{1-C\Lambda(u)} - \frac{C\hat{q}}{2} - \frac{1}{2}\mathbf{m}^{\mathsf{T}}\mathbf{K}\mathbf{m}\\ &+ \left\langle \frac{2a}{\sqrt{2\pi}}e^{-\frac{b^2}{2a^2}} + b\operatorname{erf}\left(\frac{b}{\sqrt{2}a}\right)\right\rangle,\end{aligned} \tag{D6}$$

where $a = \sqrt{\hat{q}}$ and $b = \boldsymbol{\xi}_F^{\mathsf{T}}\mathbf{K}\mathbf{m}$.

1. The spin-glass solution

In the spin-glass solution of the SDE, $m_\mu = 0$ for all $\mu = 1, 2, \dots, S$. Hence, we have

$$C = \sqrt{\frac{2}{\pi\hat{q}}} \tag{D7}$$

and

$$\hat{q} = \alpha\int_0^1 du\frac{\Lambda^2(u)}{[1-C\Lambda(u)]^2}. \tag{D8}$$

We consider the simplest case of $\gamma = 0$ and $c = 1$. It immediately follows that

$$\hat{q} = \frac{\alpha}{(1-C)^2}. \tag{D9}$$

Therefore, $C = (1 + \sqrt{\frac{\pi\alpha}{2}})^{-1}$, recovering previous results in the Hopfield model.

2. The retrieval solution

The ferromagnetic phase has a single nonvanishing overlap, i.e., $m_\mu = m\delta_{\mu,1} \sim O(1)$. They are called retrieval states, captured by the following equations:

$$m = \left\langle \xi^1 \operatorname{erf}\left[\frac{1}{\sqrt{2\hat{q}}}m\left[\boldsymbol{\xi}_F^{\mathsf{T}}\mathbf{K}\right]_1\right]\right\rangle, \tag{D10a}$$

$$C = \sqrt{\frac{2}{\pi\hat{q}}}\left\langle \exp\left[-\frac{\left[m\left[\boldsymbol{\xi}_F^{\mathsf{T}}\mathbf{K}\right]_1\right]^2}{2\hat{q}}\right]\right\rangle, \tag{D10b}$$

$$\hat{q} = \alpha\int_0^1 du\frac{\Lambda^2(u)}{[1-C\Lambda(u)]^2} + \left[\frac{\partial \mathbf{K}}{\partial C}\right]_{11}m^2. \tag{D10c}$$

In the simplest case of $\gamma = 0$ and $c = 1$, we have $\mathbf{K} = \mathbb{1}$. The above equations thus reduce to

$$m = \operatorname{erf}\left(\frac{m}{\sqrt{2\hat{q}}}\right), \tag{D11a}$$

$$C = \sqrt{\frac{2}{\pi\hat{q}}}e^{-\frac{m^2}{2\hat{q}}}, \tag{D11b}$$

$$\hat{q} = \frac{\alpha}{(1-C)^2}. \tag{D11c}$$

This result gives the memory capacity of $\alpha_c \simeq 0.138$, beyond which $\mathbf{m} = 0$, which is exactly the memory capacity of the standard Hopfield network [19]. In the general case we consider in this paper, it is necessary to solve the general equation numerically.

APPENDIX E: MEAN-FIELD DYNAMICS OF THE MODEL

In this section, we give a detailed derivation of the mean-field iterative equation for the overlap, and the formula for computing the correlation between two stimulus-induced attractors. These attractors are the neural activity fixed points when a specified pattern is initialized to a zero-temperature dynamics of the model.

The overlap describes the similarity between the state of the network at the time step t and the stored pattern μ, defined by

$$m_\mu^t = \frac{1}{N}\sum_{i=1}^{N} \xi_i^\mu s_i^t. \tag{E1}$$

The update rule of the network is given by

$$s_i^{t+1} = \mathrm{sgn}\left(\sum_j J_{ij} s_j^t\right). \tag{E2}$$

Hence,

$$\frac{1}{N}\sum_i \xi_i^\mu s_i^{t+1} = \frac{1}{N}\sum_i \xi_i^\mu\, \mathrm{sgn}\left(\sum_j J_{ij} s_j^t\right). \tag{E3}$$

Inserting the explicit expression of $\mathbf{J}$ into Eq. (E3), we arrive at

$$\frac{1}{N}\sum_i \xi_i^\mu s_i^{t+1} = \frac{1}{N}\sum_i \xi_i^\mu\, \mathrm{sgn}\left[\frac{1}{N}\sum_j\sum_\mu\left(c\xi_i^\mu\xi_j^\mu + \gamma\sum_{r=1}^{d}\left(\xi_i^{\mu+r}\xi_j^\mu + \xi_i^\mu\xi_j^{\mu+r}\right)\right)s_j^t\right]. \tag{E4}$$

Using the definition of the overlap and the cyclic feature of the pattern sequence, we have

$$m_\mu^{t+1} = \frac{1}{N}\sum_i \xi_i^\mu\, \mathrm{sgn}\left[\sum_{\mu=1}^{P} m_\mu^t\left(c\xi_i^\mu + \gamma\sum_{r=1}^{d}\left(\xi_i^{\mu+r} + \xi_i^{\mu-r}\right)\right)\right]. \tag{E5}$$

Thus the change Δm_μ caused by the update is given by

$$\Delta m_\mu = \frac{1}{N}\sum_i \xi_i^\mu\, \mathrm{sgn}\left[\sum_{\mu=1}^{P} m_\mu^t\left(c\xi_i^\mu + \gamma\sum_{r=1}^{d}\left(\xi_i^{\mu+r} + \xi_i^{\mu-r}\right)\right)\right] - m_\mu^t. \tag{E6}$$

When the network size is large enough (but $P/N \to 0$), we apply the mean-field approximation, i.e., the behavior of Eq. (E6) converges to the typical behavior of the same quantities averaged over the quenched disorder of stored patterns. More precisely, we have

$$\Delta m_\mu = \left\langle \xi^\mu\, \mathrm{sgn}\left[\sum_{\mu=1}^{P} m_\mu^t\left(c\xi^\mu + \gamma\sum_{r=1}^{d}(\xi^{\mu+r} + \xi^{\mu-r})\right)\right]\right\rangle_{\boldsymbol{\xi}} - m_\mu^t, \tag{E7}$$

where $\langle\cdot\rangle$ denotes the average over $\{\xi^1, \xi^2, \dots, \xi^P\}$. Δm_μ must vanish when a stationary solution is arrived. Therefore,

$$m_\mu = \left\langle \xi^\mu\, \mathrm{sgn}\left[\sum_{\mu=1}^{P} m_\mu\left(c\xi^\mu + \gamma\sum_{r=1}^{d}(\xi^{\mu+r} + \xi^{\mu-r})\right)\right]\right\rangle_{\boldsymbol{\xi}}, \tag{E8}$$

where we have assigned the pattern index to the superscript. By making a pattern-index shift and using the property of the cyclic sequence, we recast Eq. (E8) into the following form:

$$m_\mu = \left\langle \xi^\mu\, \mathrm{sgn}\left[\sum_{\mu=1}^{P} \xi^\mu\left(cm_\mu + \gamma\sum_{r=1}^{d}\left(m_{\mu+r} + m_{\mu-r}\right)\right)\right]\right\rangle_{\boldsymbol{\xi}}. \tag{E9}$$

The vector-form version of Eq. (E9) is given by

$$\mathbf{m} = \langle \boldsymbol{\xi} \times \mathrm{sgn}(\boldsymbol{\xi}\cdot\tilde{\mathbf{m}})\rangle_{\boldsymbol{\xi}}, \tag{E10}$$

where $\tilde{\mathbf{m}} = c\mathbf{m} + \gamma\sum_{r=1}^{d}(\mathbf{m}^{\to r} + \boldsymbol{m}^{\leftarrow r})$. $\mathbf{m}^{\to r}$ denotes a transformed $\mathbf{m}$ by shifting forward the original $\mathbf{m}$ by r patterns, while $\mathbf{m}^{\leftarrow r}$ denotes shifting backward the original $\mathbf{m}$ by r patterns.

Algorithm 1. Procedure to calculate the overlap from Eq. (E10)

```
Input: The number of the patterns $P$; The number of the Monte Carlo samples $\mathcal{T}$; Network parameters $c, d, \gamma$; Convergence precision $\epsilon$;
    Iteration rate $\eta$.
Output: The overlap vector $\boldsymbol{m} \equiv \{m_1, m_2, \dots, m_P\}$.
Sample $\mathcal{T}$ $P$-length stored patterns $\{\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, \cdots \boldsymbol{\xi}^{\mathcal{T}}\}$.
Initialize the overlap vector as $\boldsymbol{m} \leftarrow [0, \dots, 0, 1, 0, \dots, 0]$, where the value of 1 means the overlap with the stimulating pattern.
Initialize an intermediate vector $\boldsymbol{m}' \leftarrow [0, \dots, 0, 0, 0, \dots, 0]$.
while$\|\boldsymbol{m} - \boldsymbol{m}'\|_2^2 > \epsilon$
    $\boldsymbol{m}' \leftarrow \boldsymbol{m}$
    $\tilde{\boldsymbol{m}} \leftarrow c\boldsymbol{m} + \gamma \sum_{r=1}^{d} (\boldsymbol{m}^{\rightarrow r} + \boldsymbol{m}^{\leftarrow r})$
    Initialize rhs $\leftarrow [0, \dots, 0]$
    for $i = 1 \rightarrow \mathcal{T}$
        rhs $\leftarrow$ rhs $+ \frac{1}{\mathcal{T}} \boldsymbol{\xi}^i \cdot \mathrm{sgn}(\boldsymbol{\xi}^i \cdot \tilde{\mathbf{m}})$
    end for
    $\mathbf{m} \leftarrow \eta \times \mathbf{m}' + (1 - \eta) \times$ rhs.
end while
return m
```

In an analogous way, the correlation between stimulus-induced attractors is given by

$$\mathcal{C}(\alpha, \beta) = \left\langle \mathrm{sgn}\left\{ \sum_{\mu=1}^{P} \xi^{\mu} \left[c m_{\mu}^{\alpha} + \gamma \sum_{i=1}^{d} \left(m_{\mu+i}^{\alpha} + m_{\mu-i}^{\alpha} \right) \right] \right\} \mathrm{sgn}\left\{ \sum_{\mu=1}^{P} \xi^{\mu} \left[c m_{\mu}^{\beta} + \gamma \sum_{i=1}^{d} \left(m_{\mu+i}^{\beta} + m_{\mu-i}^{\beta} \right) \right] \right\} \right\rangle_{\boldsymbol{\xi}}, \tag{E11}$$

where m_{μ}^{α} means the μth overlap when the system lies in the attractor induced by the pattern α, i.e., the solution of Eq. (E9) initialized with $m_{\mu} = \delta_{\mu\alpha}$. Because of the structure of the attractors, the correlation depends only on the separation of the corresponding stimulating patterns in the stored cyclic sequence. Therefore, we rewrite Eq. (E11) as

$$\mathcal{C}(r) = \left\langle \mathrm{sgn}\left[\sum_{\mu=1}^{P} \xi^{\mu} \left(c m_{\mu} + \gamma \sum_{i=1}^{d} (m_{\mu+i} + m_{\mu-i}) \right) \right] \mathrm{sgn}\left[\sum_{\nu=1}^{P} \xi^{\nu} \left(c m_{\nu}^{\rightarrow r} + \gamma \sum_{i=1}^{d} \left(m_{\nu+i}^{\rightarrow r} + m_{\nu-i}^{\rightarrow r} \right) \right) \right] \right\rangle_{\boldsymbol{\xi}}. \tag{E12}$$

A pseudocode to solve Eq. (E9) is shown in the Algorithm 1. In the algorithm, $\mathcal{T}$ denotes the the number of Monte Carlo samples. In practice, we set $\mathcal{T} = 5 \times 10^5$ in a single trial, and the result is averaged over 30 trials. The correlation between attractors can be estimated from the solution of the overlap (see Algorithm 2).

Algorithm 2. Procedure to calculate the attractor correlation from Eq. (E11)

```
Input: The number of the patterns $P$; The number of the Monte Carlo samples $\mathcal{T}$; Network parameters $c, d, \gamma$;  calculated from Algorithm 1;
    Random patterns $\{\boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^{\mathcal{T}}\}$.
Output: The correlation as a function of different separations $\boldsymbol{C} = [\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_{P/2}]$.
Initialize $\boldsymbol{C} = [0, \dots, 0]$
$\tilde{\mathbf{m}} \leftarrow c\mathbf{m} + \gamma \sum_{r=1}^{d} (\mathbf{m}^{\rightarrow r} + \mathbf{m}^{\leftarrow r})$
for $i = 1 \rightarrow P/2$
    for $j = 1 \rightarrow \mathcal{T}$
        $s_1 \leftarrow \mathrm{sgn}(\tilde{\boldsymbol{m}} \cdot \boldsymbol{\xi}^j)$
        $s_2 \leftarrow \mathrm{sgn}(\tilde{\boldsymbol{m}}^{\rightarrow i} \cdot \boldsymbol{\xi}^j)$
        $\mathcal{C}_i \leftarrow \mathcal{C}_i + \frac{1}{\mathcal{T}} s_1 \cdot s_2$
    end for
end for
return $\boldsymbol{C}$
```

[1] S. J. Guzman, A. Schlogl, M. Frotscher, and P. Jonas, Synaptic mechanisms of pattern completion in the hippocampal CA3 network, Science **353**, 1117 (2016).

[2] M. S. Ahmed, J. B. Priestley, A. Castro, F. Stefanini, A. S. S. Canales, E. M. Balough, E. Lavoie, L. Mazzucato, S. Fusi, and A. Losonczy, Hippocampal network reorganization underlies

the formation of a temporal association memory, Neuron **107**, 283 (2020).

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

[4] S.-i. Amari, Neural theory of association and concept-formation, Biol. Cybern. **26**, 175 (1977).

[5] Y. Miyashita, Neuronal correlate of visual associative long-term memory in the primate temporal cortex, Nature (London) **335**, 817 (1988).

[6] Y. Miyashita and H. Chang, Neuronal correlate of pictorial short-term memory in the primate temporal cortex, Nature (London) **331**, 68 (1988).

[7] M. Griniasty, M. V. Tsodyks, and D. J. Amit, Conversion of temporal correlations between stimuli to spatial correlations between attractors, Neural Comput. **5**, 1 (1993).

[8] T. Haga and T. Fukai, Extended Temporal Association Memory by Modulations of Inhibitory Circuits, Phys. Rev. Lett. **123**, 078101 (2019).

[9] K. C. Bittner, A. D. Milstein, C. Grienberger, S. Romani, and J. C. Magee, Behavioral time scale synaptic plasticity underlies CA1 place fields, Science **357**, 1033 (2017).

[10] W. Gerstner, M. Lehmann, V. Liakoni, D. Corneil, and J. Brea, Eligibility traces and plasticity on behavioral time scales: Experimental support of neoHebbian three-factor learning rules, Front. Neural Circuits **12**, 53 (2018).

[11] E. T. Reifenstein, I. B. Khalid, and R. Kempter, Synaptic learning rules for sequence learning, eLife **10**, E67171 (2021).

[12] M. E. Hasselmo, Neuromodulation: Acetylcholine and memory consolidation, Trends Cogn. Sci. **3**, 351 (1999).

[13] F. Crick and G. Mitchison, The function of dream sleep, Nature (London) **304**, 111 (1983).

[14] J. J. Hopfield, D. I. Feinstein, and R. G. Palmer, "Unlearning" has a stabilizing effect in collective memories, Nature (London) **304**, 158 (1983).

[15] S. Diekelmann and J. Born, The memory function of sleep, Nat. Rev. Neurosci. **11**, 114 (2010).

[16] Y. Zhou, C. S. W. Lai, Y. Bai, W. Li, R. Zhao, G. Yang, M. G. Frank, and W.-B. Gan, REM sleep promotes experience-dependent dendritic spine elimination in the mouse cortex, Nat. Commun. **11**, 4819 (2020).

[17] D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).

[18] L. F. Cugliandolo and M. V. Tsodyks, Capacity of networks with correlated attractors, J. Phys. A: Math. Gen. **27**, 741 (1994).

[19] 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).

[20] R. Kempter, W. Gerstner, and J. L. van Hemmen, Hebbian learning and spiking neurons, Phys. Rev. E **59**, 4498 (1999).

[21] L. Abbott and S. Nelson, Synaptic plasticity: Taming the beast, Nat. Neurosci. **3**, 1178 (2000).

[22] J.-P. Pfister and W. Gerstner, Triplets of spikes in a model of spike timing-dependent plasticity, J. Neurosci. **26**, 9673 (2006).

[23] R. M. Gray, Toeplitz and circulant matrices: A review, Found. Trends Commun. Inf. Theory **2**, 155 (2006).

[24] T. Hou, K. Y. M. Wong, and H. Huang, Minimal model of permutation symmetry in unsupervised learning, J. Phys. A: Math. Theor. **52**, 414001 (2019).

[25] T. Hou and H. Huang, Statistical Physics of Unsupervised Learning with Prior Knowledge in Neural Networks, Phys. Rev. Lett. **124**, 248302 (2020).

[26] J. Zhou, Z. Jiang, T. Hou, Z. Chen, K. Y. M. Wong, and H. Huang, Eigenvalue spectrum of neural networks with arbitrary Hebbian length, Phys. Rev. E **104**, 064307 (2021).

[27] G. R. Poe, Sleep is for forgetting, J. Neurosci. **37**, 464 (2017).

[28] D. Kleinfeld and D. B. Pendergraft, Unlearning increases the storage capacity of content addressable memories, Biophys. J. **51**, 47 (1987).

[29] V. S. Dotsenko, N. D. Yarunin, and E. A. Dorotheyev, Statistical mechanics of Hopfield-like neural networks with modified interactions, J. Phys. A **24**, 2419 (1991).

[30] K. Nokura, Spin glass states of the anti-Hopfield model, J. Phys. A: Math. Gen **31**, 7447 (1998).

[31] A. Fachechi, E. Agliari, and A. Barra, Dreaming neural networks: Forgetting spurious memories and reinforcing pure ones, Neural Netw. **112**, 24 (2019).

[32] C. C. A. Fung, K. Y. M. Wong, and S. Wu, A moving bump in a continuous manifold: A comprehensive study of the tracking dynamics of continuous attractor neural networks, Neural Comput. **22**, 752 (2010).

[33] A. Battista and R. Monasson, Capacity-Resolution Trade-Off in the Optimal Learning of Multiple Low-Dimensional Manifolds by Attractor Neural Networks, Phys. Rev. Lett. **124**, 048302 (2020).

[34] G. Mongillo, D. J. Amit, and N. Brunel, Retrospective and prospective persistent activity induced by Hebbian learning in a recurrent cortical network, Eur. J. Neurosci. **18**, 2011 (2003).

[35] N. Brunel and F. Lavigne, Semantic priming in a cortical network model, J. Cogn. Neurosci. **21**, 2300 (2009).

[36] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).